Indira Gandhi National Open University
School of Sciences

MST-012
PROBABILITY AND PROBABILITY DISTRIBUTIONS

Block 2

DISCRETE PROBABILITY DISTRIBUTIONS

BLOCK 1: Probability Measure and Random Variable

Unit 1: Review of Basic Concepts of Probability

Unit 2: Probability Space

Unit 3: Assignment of Probabilities to Events in Discrete and Continuous Worlds

Unit 4: Univariate Discrete Random Variable

Unit 5: Univariate Continuous Random Variable

Unit 6: Bivariate Random Variable

Unit 7: Expectation and Moment Generating Function

Unit 8: Transformation of Univariate and Bivariate Random Variable

BLOCK 2: Discrete Probability Distributions

Unit 9: Uniform and Bernoulli Distributions

Unit 10: Binomial and Multinomial Distributions

Unit 11: Poisson and Hypergeometric Distributions

Unit 12: Geometric and Negative Binomial Distributions

BLOCK 3: Continuous Probability Distributions

Unit 13: Uniform and Exponential Distributions

Unit 14: Normal and Lognormal Distributions

Unit 15: Gamma and Beta Distributions

Unit 16: Laplace and Cauchy Distributions

BLOCK 4: Inequalities and Sequence of Random Variables

Unit 17: Inequalities

Unit 18: Convergence of Sequence of Random Variables

Unit 19: Laws of Large Numbers

Unit 20: Central Limit Theorem

BLOCK 2: DISCRETE PROBABILITY DISTRIBUTIONS

In Block 1, you have studied random variables and the CDF of a random variable. In the same block, you have also studied the PMF of a discrete random variable and the PDF of a continuous random variable. In the present block, we will study discrete probability distributions so we will use the idea of PMF and CDF very frequently. In Block 1, you have also learnt to obtain summary measures of a probability distribution of a random variable. We will also use these ideas in this block very frequently. The present block contains four Units 9 to 12. This block discusses some well-known discrete probability distributions.

Unit 9: This unit discusses summary measures and applications of Bernoulli and discrete uniform distributions.

Unit 10: This unit discusses summary measures and applications of binomial and multinomial distributions.

Unit 11: This unit discusses summary measures and applications of Poisson and hypergeometric distributions.

Unit 12: This unit discusses summary measures and applications of geometric and negative binomial distributions.

Expected Learning Outcomes

After completing this block, you should be able to:

- apply Bernoulli, binomial and multinomial distributions to solve problems based on them; and
- apply discrete uniform, Poisson, hypergeometric, geometric and negative binomial distributions to solve problems based on them.

Block Preparation Team

Notations and Symbols

Sec./Secs.	:	Section/Sections
Fig./Figs.	:	Figure/Figures
$X \sim \text{dunif}(a, b)$	:	X follows discrete uniform distribution with parameters a and b
$X \sim \text{Bern}(p)$	:	X follows Bernoulli distribution with parameter p
$X \sim \text{Bin}(n, p)$	:	X follows binomial distribution with parameters n and p
$X \sim \text{Pois}(\lambda)$	:	X follows Poisson distribution with parameter λ
$X \sim \text{HGeom}(G, B, n)$	:	X follows hypergeometric distribution with parameters G, B and n
$X \sim \text{Geom}(p)$	:	X follows geometric distribution with parameter p
$X \sim \text{NBin}(r, p)$	:	X follows negative binomial distribution with parameters r and p

UNIT 9

UNIFORM AND BERNOULLI DISTRIBUTIONS

Structure

9.1 INTRODUCTION

In Unit 4, you have studied the following concepts that we are going to use in the present unit as well as in the next three units:

- definition of random variable you may refer to 4.6 and 4.14) and in particular definition of discrete random variable you may refer to 4.57 and 4.58);
- cumulative distribution function (CDF) in general, you may refer to 4.35 and 4.36 as well as the properties of CDF you may refer to 4.52 to 4.55. In particular, the definition of CDF for the discrete random variable has also been discussed you may refer to 4.63 and 4.64; and
- probability mass function (PMF) of a discrete random variable you may refer to 4.59 and 4.60.

In the present unit, we will use these concepts for two particular discrete probability distributions namely discrete uniform and Bernoulli distributions. In Sec. 9.2, you will get to know what are the requirements that should be fulfilled to apply discrete uniform distribution. In the same section, you will also study PMF and CDF of the discrete uniform distribution. Moment generating function (MGF) and some other summary measures like mean, variance, etc. will be discussed for the same probability distribution in Sec. 9.3. In Secs. 9.4 and 9.5, we will do similar studies for Bernoulli distribution. Some applications and analysis of some measures of these distributions are discussed in Sec. 9.6.

What we have discussed in this unit is summarised in Sec. 9.7. Self-Assessment Questions (SAQs) have been given in some sections which are generally based on the content discussed in that section. But to give you a good practice of what we have discussed in this unit some more questions based on the entire unit are given in Sec. 9.8 under the heading Terminal Questions. Due to the reason mentioned in Sec. 1.1 of Unit 1 of this course, solutions of all the SAQs and Terminal Questions are given in Sec. 9.9.

In the next unit, you will do a similar study about two more discrete probability distributions known as binomial and multinomial distributions.

Expected Learning Outcomes

After completing this unit, you should be able to:

- ❖ identify situations where you can apply discrete uniform and Bernoulli distributions;
- ❖ define PMF, CDF, MGF and some summary measures of discrete uniform and Bernoulli distributions; and
- ❖ apply discrete uniform and Bernoulli distributions to solve problems based on these two probability distributions.

9.2 STORY, DEFINITION, PMF AND CDF OF DISCRETE UNIFORM DISTRIBUTION

In Unit 4, you have studied when a random variable is said to be a discrete random variable and also studied that each discrete random variable has its PMF. You have also obtained PMF and CDF of some discrete random variables in Unit 4. In this section, we will discuss one special discrete random variable known as a discrete uniform random variable. We will also discuss the PMF and CDF of the discrete uniform distribution. Let us start our discussion with the story of the discrete uniform distribution.

Let us consider three random variables and their PMF's.

Random Variable I and its PMF: Let X denote the number of heads when a fair coin is tossed once. PMF of X for this random experiment is given by

$$p_X(x) = \begin{cases} 1/2, & \text{if } x = 0, 1 \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9.1)$$

Or in tabular form, it can be written as shown in Table 9.1 given as follows.

Table 9.1: Probability mass function or probability distribution of rv X

X	0	1
$p_X(x)$	1/2	1/2

Random Variable II and its PMF: Let Y denote the number written on the face that turns up when a fair tetrahedral die is thrown once. PMF of Y for this random experiment is given by

$$p_Y(y) = \begin{cases} 1/4, & \text{if } y = 1, 2, 3, 4 \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9.2)$$

Or in tabular form, it can be written as shown in Table 9.2 given as follows.

Table 9.2: Probability mass function or probability distribution of rv Y

Y	1	2	3	4
$p_Y(y)$	1/4	1/4	1/4	1/4

Random Variable III and its PMF: Consider a fair die having 10 faces with numbers 1 to 10 on its 10 faces. Let Z denote the number written on the face that turns up when this fair die is thrown once. PMF of Z for this random experiment is given by

$$p_Z(z) = \begin{cases} 1/10, & \text{if } z = 1, 2, 3, 4, \ldots, 10 \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9.3)$$

Or in tabular form, it can be written as shown in Table 9.3 given as follows.

Table 9.3: Probability mass function or probability distribution of rv Z

Z	1	2	3	4	5	6	7	8	9	10
$p_Z(z)$	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10	1/10

You were thinking about why we are considering three random variables and their PMFs to explain the story of the uniform distribution. We are very close to the answer to your question. Let me ask you a different question whose answer is directly related to the answer to the question which is arising in your mind. My question is keeping our target in view what are the two common things in these three PMFs of three different random variables X, Y and Z in which we may be interested. Before going further try to get an answer to this question no matter whether your answer matches with my answer or not. At least give a try. Now, compare your answer with the one I was expecting. Two common things in these three PMFs which are of our interest keeping the title of this section in view are mentioned as follows.

- The first common thing is all three random variables X, Y and Z assume a finite number of values. Here X assumes only two values 0 and 1; Y assumes only four values 1 to 4 and Z assumes only ten values 1 to 10. … (9.4)
- The second common thing is all the possible outcomes in each experiment are equally likely. If you want to recall what we mean by equally likely then refer to (1.7) of this course. … (9.5)

If your answer matches then good if not then very good. You should be happy that you gave it a try. Always remember whenever we try, we definitely learn something. Now, let us continue the story of discrete uniform distribution. Let us visualise these three PMFs of the random variables X, Y and Z in Fig. 9.1 (a), (b) and (c) respectively as follows. Note that to visualise this information, we have used one principle of data visualisation which states that if we want to compare the vertical height of two or more distributions then we have to plot graphs horizontally side by side adjacent to each other keeping the same scale on the vertical axis in each figure. Also, refer to Fig. 8.2 in the previous unit to see another similar principle of data visualisation. … (9.6)

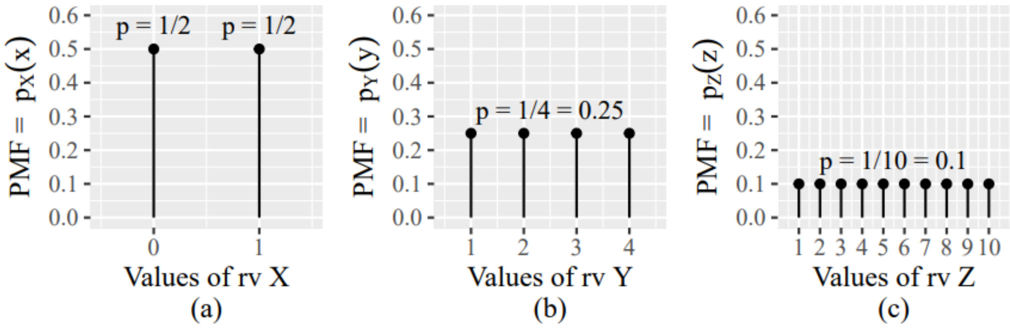


Fig. 9.1: Visualisation of PMF's of random variables (a) X (b) Y (c) Z

From Fig. 9.1 (a) note that the meaning of equally likely outcomes mentioned in (9.5) is that the probability of each value of the random variable is equal. In Fig. 9.1 (a) it is 1/2 for each value of X. In Fig. 9.1 (b) it is 1/4 for each value of Y instead of 1/2, while in Fig. 9.1 (c) it is 1/10 for each value of Z instead of 1/2 or 1/4. Another thing that should be noted from Fig. 9.1 is that when outcomes of the random experiments are equally likely then the probability of equally likely outcomes decreases as the number of outcomes increases. It is simply because the total probability is always 1, we cannot change it. So, if we distribute it equally in the more number of outcomes then definitely each outcome will get less probability compared to the experiment where outcomes are less in numbers.

Let us discuss the role of the issue of increasing the number of outcomes mathematically. Mathematically, as the number of outcomes in the random experiment of equally likely outcomes tends to ∞ then the probability of each outcome tends to 0. **So, we cannot have a random experiment having an infinite number of outcomes and all of them are equally likely**. To explain it further suppose such a random experiment exists. Then the probability of each outcome will be some fixed non-zero number ε (say), where $0 < \varepsilon < 1$. Now, the sum of all probabilities will be $\varepsilon + \varepsilon + \varepsilon + \varepsilon + \ldots$ Whatever small value of ε, we take sum of infinite such number will be greater than 1. For example,

- if $\varepsilon = 0.1$, then the sum of only 11 such numbers will be 1.1 > 1.
- if $\varepsilon = 0.01$, then the sum of only 101 such numbers will be 1.01 > 1.
- if $\varepsilon = 0.001$, then the sum of only 1001 such numbers will be 1.001 > 1.
- if $\varepsilon = 0.0001$, then the sum of only 10001 such numbers will be 1.0001 > 1.

and so on. In fact, the sum $\varepsilon + \varepsilon + \varepsilon + \varepsilon + \ldots$ will be infinite. Both the points mentioned in (9.4) and (9.5) have been explained in detail which are base of the discrete uniform random variable. So, now, we can write the story of discrete uniform distribution as follows.

Story of Discrete Uniform Distribution: If the number of possible outcomes of a random experiment satisfies the following three conditions:

- finite in number, e.g., may be 1 or 2 or 3 or 4 or 5, etc. … (9.7)

- equally spaced, e.g., 1, 2, 3, … or 1.2, 1.7, 2.2, 2.7, …, etc. But mostly we have consecutive integers instead of something like 1.2, 1.7, 2.2, 2.7, … and … (9.8)
- equally likely, … (9.9)

then we say that it is a perfect situation for the discrete uniform distribution. … (9.10)

Now, before defining discrete uniform distribution let us first give you a big picture which is shared by all discrete probability distributions. Recall from the discussion of Unit 2 (you may refer to 2.29 of Unit 2 of this course) that in the world of probability theory, we have a triplet $(\Omega, \mathcal{F}, \mathcal{P})$ where … (9.11)

- The full space Ω is a set which contains all possible outcomes of the random experiment. … (9.12)
- The second thing of this universe is $\mathcal{F}$ which is a collection of subsets of Ω which are of our interest and having at least two members ϕ and Ω. It is closed under complement and countable union. … (9.13)
- The third thing of this universe is $\mathcal{P}$, a probability measure which assigns a probability to each member of $\mathcal{F}$. … (9.14)

In Units 9 to 12, we will discuss some well-known discrete probability distributions. As we explained in Unit 3 in detail that in all discrete probability distributions $\mathcal{F}$ is always the power set of Ω. So, out of three things $\Omega, \mathcal{F}, \mathcal{P}$ one thing $\mathcal{F}$ has been fixed for all discrete probability distributions. The remaining two things Ω and $\mathcal{P}$ will vary from distribution to distribution. So, moral of the story is as soon as we specify Ω and $\mathcal{P}$ probability distribution will automatically specify. Keep this important point in mind. So, by definition of a discrete probability distribution, we mean the specification of Ω and $\mathcal{P}$. … (9.15)

Other than this big picture recall (4.59) and note that to obtain the probability distribution of our interest, we first need to define a random variable on the sample space. But we can define many random variables on one sample space refer to (4.75). We also know that each random variable has its own PMF or PDF and CDF. So, in what type of probability distribution, we are interested accordingly, we define a random variable and obtain PMF or PDF or CDF of the random variable and do analysis as per the problem in hand. So, ideally, to obtain a probability distribution of our interest first we should define sample space Ω of the random experiment then a random variable on Ω and then PMF or PDF or CDF with the help of probability measure $\mathcal{P}$ already defined on the sample space Ω. Once, we get PMF or PDF or CDF it means we have a probability distribution of our interest at our disposal to do analysis according to the question we want to address. Keeping all this in view let us define the first probability distribution of this course.

Now, we define discrete uniform probability distribution as follows.

Definition and PMF of Discrete Uniform Distribution: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \{a, a+1, a+2, a+3, \ldots, b-1, b\}$ contains $b - a + 1 = N$ (say), number of outcomes of a random experiment which are finite in number as well as equally likely. Let X be a random variable defined on the sample space Ω by $X(\omega) = \omega \ \forall \ \omega \in \Omega$. So, random variable X assumes N equally spaced values $a, a+1, a+2, \ldots, b-1, b$. We say that the random variable X follows discrete uniform distribution if probability measure $\mathcal{P}$ assigned equal probability to each value of X, i.e., if

$$\mathcal{P}(X = x) = \frac{1}{b-a+1}, \ x = a, a+1, a+2, a+3, \ldots, b \qquad \ldots (9.16)$$

where $b - a + 1 = N$ … (9.17)

So, PMF of discrete uniform random variable X is given by

$$\mathcal{P}(X = x) = \begin{cases} \dfrac{1}{b-a+1}, & x = x = a, a+1, a+2, a+3, \ldots, b \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9.18)$$

where $X = x$ is the event E where $E = \{\omega \in \Omega : X(\omega) = x\}$

So,

$X = a = a + 0,$ is the event E_0 where $E_0 = \{\omega \in \Omega : X(\omega) = a\} = \{a\}$

$X = a + 1,$ is the event E_1 where $E_1 = \{\omega \in \Omega : X(\omega) = a + 1\} = \{a + 1\}$

$X = a + 2,$ is the event E_2 where $E_2 = \{\omega \in \Omega : X(\omega) = a + 2\} = \{a + 2\}$

$\vdots$

$X = a + N - 1 = b,$ is the event E_{N-1} where $E_{N-1} = \{\omega \in \Omega : X(\omega) = b\} = \{b\}$

So,

$$\mathcal{P}(X = a + k) = \mathcal{P}(E_k) = \mathcal{P}(\{\omega \in \Omega : X(\omega) = a + k\})$$

$$= \mathcal{P}(\{a + k\}) = \frac{1}{b-a+1} \quad \left[\begin{array}{l} \because n(\{a+k\}) = 1, \ n(\Omega) = b - a + 1 = N \text{ and} \\ \text{using uniform probability law. You may} \\ \text{refer to (3.7) in Unit 3 of this course} \end{array}\right]$$

If random variable X follows discrete uniform distribution and assumes N equally spaced values $a, a+1, a+2, \ldots, b-1, b,$ where $N = b - a + 1,$ then a and b are known as parameters of discrete uniform distribution and is denoted by writing $X \sim \text{DUnif}(a, b)$. … (9.19)

The symbol '~' is read as 'is distributed as' or 'follows'. So, $X \sim \text{DUnif}(a, b)$ is read as X is distributed as discrete uniform distribution with parameters a and b. Or we read it as X follows discrete uniform distribution with parameters a and b. … (9.20)

Since the statistical software used for hands on practice for the lab courses of this programme is R programming language, so you should also know its notation in R. In R, notation that is used for discrete uniform distribution is dunif(min, max), where min and max represents values of a and b respectively. In fact, like any probability distribution there are four functions for discrete uniform probability distribution namely ddunif(x, min, max, …), pdunif(q, min, max, …), qdunif(p, min, max, …) and rdunif(n, min, max, …). … (9.21)

Role of each of these four functions is the same for all probability distributions explained as follows.

- First d in ddunif() represents density in the case of continuous random variable and PMF in the case of discrete random variable and it gives $\text{ddunif}(\text{x, min, max}) = \mathcal{P}(X = x).$ … (9.22)

- First letter p in pdunif() represents distribution function or CDF and it gives $\text{pdunif}(\text{q, min, max}) = \mathcal{P}(X \le q).$ … (9.23)

- First letter q in qdunif() represents quantile function and it gives value of p such that $\mathcal{P}(X \le p) = q.$ … (9.24)

- First letter r in rdunif() represents random number generation and $\text{rdunif}(\text{n, min, max})$ generates n random numbers from discrete uniform with parameters min and max. … (9.25)

Let us check the **validity of the PMF of the discrete uniform distribution**.

We have to check two conditions non-negativity and normality refer to (4.62).

(1) **Non-negativity**: Since $b \ge a \Rightarrow b - a \ge 0 \Rightarrow b - a + 1 > 0 \Rightarrow \frac{1}{b-a+1} > 0$ so

$$\mathcal{P}(X = x) = \frac{1}{b-a+1} > 0 \quad \forall \quad x = a, a+1, a+2, a+3, \ldots, b \qquad \text{… (9.26)}$$

(2) **Normality**: Number of terms $a, a+1, a+2, a+3, \ldots, b$ are b – a + 1. So,

$$\sum_{k=a}^{b} \frac{1}{b-a+1} = (b-a+1)\frac{1}{b-a+1} \quad \left[\because \sum_{k=1}^{n} a = na \text{ if } a \text{ is independent of } k\right]$$

$$= 1 \qquad \text{... (9.27)}$$

This proves that sum of all probabilities of discrete uniform distribution is 1.

Hence, we can say that PMF given by (9.18) is a valid PMF.

One question that will be arising in your mind is how the expression $\frac{1}{b-a+1}$ comes in the definition of probability measure given by (9.16) or (9.18) or how we are saying that there are b – a + 1 terms in a, a + 1, a + 2, a + 3, …, b. Let us explain it. Here, random variable X takes values a, a + 1, a + 2, a + 3, …, b. Here a and b are integers. We know that if we count the numbers:

- 1, 2, 3, 4, …, 10 then we find that they are 10 in numbers.

- 1, 2, 3, 4, …, 17 then we find that they are 17 in numbers.

- 1, 2, 3, 4, …, n then we find that they are n in numbers.

- We are interested in counting the numbers of the type $k, k+1, k+2, k+3, \ldots, n$ which start from some integer k. We claim that they are n – k + 1 in numbers. Why? Let us explain it. Let us includes first k – 1 natural numbers in these numbers. After doing so, we have

  $1, 2, 3, 4, \ldots, k-1, k, k+1, k+2, k+3, \ldots, n$

  Now, these numbers are consecutive natural numbers from 1 to n so they are n in numbers. Similarly, the numbers $1, 2, 3, 4, \ldots, k-1$ are k – 1 in numbers. Hence, using three steps shown as follows:

$$\underbrace{\underbrace{1, 2, 3, 4, \ldots, k-1}_{\text{Step 2: } k-1 \text{ in numbers}}, \underbrace{k, k+1, k+2, k+3, \ldots, n}_{\text{Step 3: } n-(k-1)=n-k+1 \text{ in numbers}}}_{\text{Step 1: } n \text{ in numbers}}$$

We can say that the numbers $k, k+1, k+2, k+3, \ldots, n$ are n – k + 1 in numbers. Thus, the numbers $a, a+1, a+2, \ldots, b-1, b$ are b – a + 1 in numbers.

This completes the explanation of the answer of your question. … (9.28)

Now, we define CDF of discrete uniform distribution.

CDF of Discrete Uniform Distribution: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \{a, a+1, a+2, a+3, \ldots, b-1, b\}$ contains $b-a+1 = N$ (say), number of outcomes of a random experiment which are finite in number as well as equally likely. Let X be a random variable defined on the sample space Ω by $X(\omega) = \omega \ \forall \ \omega \in \Omega$. So, random variable X assumes N equally spaced values $a, a+1, a+2, \ldots, b-1, b.$ The random variable X follows discrete random variable if its PMF is given by

$$\mathcal{P}(X = x) = \begin{cases} \dfrac{1}{b-a+1}, & x = x = a, a+1, a+2, a+3, \ldots, b \\ 0, & \text{otherwise} \end{cases}$$

Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}(X \le x) = \sum_{k=0}^{[x]} \mathcal{P}(X = k) = \begin{cases} 0, & \text{if } x < a \\ \dfrac{[x]-a+1}{b-a+1}, & \text{if } a \le x < b \\ 1, & \text{if } x \ge b \end{cases} \quad \ldots (9.29)$$

where $[x]$ represents greatest integer function of x, to recall it you may refer to (1.54) of the course MST-011.

Let us do one example.

Example 1: A fair die is thrown once. If X denotes the number that is written on the face that turns up then name the probability distribution that X follows and why. Also, find PMF and CDF of X and plot them. Also, find values of $F_X(4.998),\ F_X(2.042),\ F_X(0.24)$ and $F_X(6.001)$.

Solution: Here random variable X may assume values 1, 2, 3, 4, 5 and 6. Since die is fair so all the values 1 to 6 are equally likely. Hence, the random variable X assumes finite number of values which are equally likely so it follows discrete uniform distribution.

PMF of the random variable X is given by

$$\mathcal{P}(X = x) = \begin{cases} \dfrac{1}{6}, & x = 1, 2, 3, 4, 5, 6 \\ 0, & \text{otherwise} \end{cases} \quad \ldots (9.30)$$

CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}(X \le x) = \begin{cases} 0, & \text{if } x < 1 \\ \dfrac{[x]}{6}, & \text{if } 1 \le x < 6 \\ 1, & \text{if } x \ge 6 \end{cases} \quad \ldots (9.31)$$

PMF and CDF of the random variable X are plotted in Fig. 9.2 (a) and (b) respectively as follows.

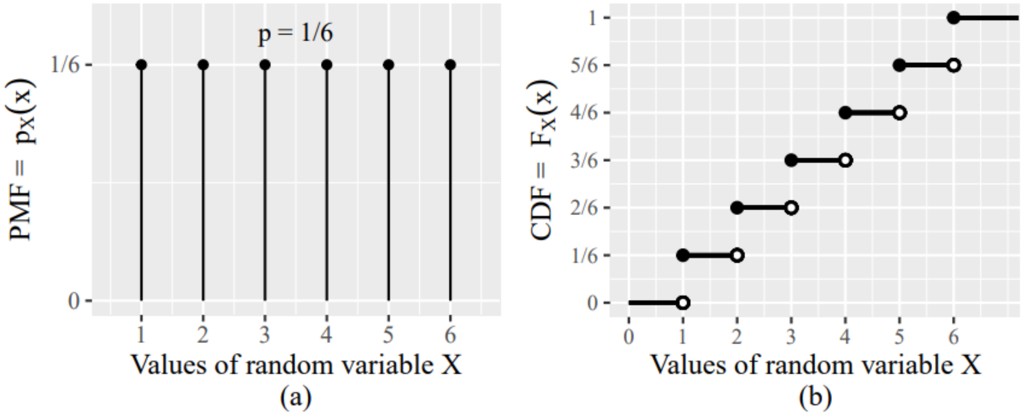


Fig. 9.2: Visualisation of (a) PMF of rv X (b) CDF of rv X of Example 1

Now, by definition of CDF refer to (9.31), we have

$$F_X(4.998) = \frac{[4.998]}{6} = \frac{4}{6} = \frac{2}{3} \quad \begin{bmatrix} \text{Why } [4.998] = 4, \text{ you may refer to} \\ \text{(1.54) of the course MST-011} \end{bmatrix}$$

$$\text{Similarly, } F_X(2.042) = \frac{[2.042]}{6} = \frac{2}{6} = \frac{1}{3}.$$

$$F_X(0.24) = 0 \ [\because 0.24 < 1] \text{ and } F_X(6.001) = 1 \ [\because 6.001 > 6].$$

Remark 1: If we have a = b then in this special case we say that X is a constant random variable having only one value ‘a’ with $\mathcal{P}(X = a) = 1$. At this point of time one question that will be arising in your mind is: “if X is taking only one value ‘a’ then it is a deterministic variable so why we are calling it a random variable”. Arising this type of questions in your mind proves that you are focusing on learning the subject like a master degree learner. Being a master degree learner, you should know the reason of each equality sign in an expression. Here, we are calling it a random variable because we are considering it as defined on sample space of a random experiment. So, probability will be associated with its value. … (9.32)

9.3 MGF AND OTHER SUMMARY MEASURES OF DISCRETE UNIFORM DISTRIBUTION

In the previous section, you have studied PMF and CDF of discrete uniform distribution. In this section we want to obtain MGF and some other summary measure of discrete uniform distribution like mean, median, variance, etc. Let us first obtain MGF of discrete uniform distribution.

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x=a}^{b} e^{tx} p_X(x) \qquad [\text{By definition of MGF refer to (7.48)}]$$

$$= \sum_{x=a}^{b} e^{tx} \frac{1}{N}, \quad \text{where } N = b - a + 1$$

$$= \frac{1}{N}\sum_{x=a}^{b} e^{tx} = \frac{1}{N}\sum_{x=a}^{b} \left(e^t\right)^x = \frac{1}{N}\left[\left(e^t\right)^a + \left(e^t\right)^{a+1} + \left(e^t\right)^{a+2} + \ldots + \left(e^t\right)^b\right]$$

$$= \frac{1}{N}\left[\frac{e^{ta}\left(1-e^{tN}\right)}{1-e^{t}}\right] \left[\begin{array}{l}\text{Sum of the first n terms of a GP} \\ a + ar + ar^2 + \ldots + ar^{n-1} = \dfrac{a\left(1-r^n\right)}{1-r} \\ \text{In our GP: } a = e^{ta},\ r = e^{t},\ n = b-a+1 = N\end{array}\right]$$

$$\Rightarrow M_X(t) = \frac{1}{N}\left[\frac{e^{ta}\left(1-e^{tN}\right)}{1-e^{t}}\right],\ t \neq 0 \qquad \ldots (9.33)$$

$$\Rightarrow M_X(t) = \frac{1}{N}\left[\frac{e^{ta}-e^{t(a+N)}}{1-e^{t}}\right],\ t \neq 0 \qquad \ldots (9.34)$$

But $N = b-a+1 \Rightarrow N + a = b + 1.$ So, we have

$$\Rightarrow M_X(t) = \frac{1}{(b-a+1)}\left[\frac{e^{at}-e^{(b+1)t}}{1-e^{t}}\right],\ t \neq 0 \qquad \ldots (9.35)$$

Recall (7.57) to (7.60) all the raw moments can be obtained by substituting t = 0 in the expressions of different derivatives of MGF. But here we cannot do that because derivatives of MGF at t = 0 are not defined. So, we have to obtain measures of our interest individually.

Mean and variance of any probability distribution are more commonly used. So, let us obtain them. But for obtaining mean and variance we need first two raw moments of discrete uniform distribution. So, let us obtain first two raw moments.

$$\text{First raw moment} = \mu_1' = E(X) = \sum_{x=a}^{b} x\,\mathcal{P}\left(X = x\right) = \sum_{x=a}^{b} x\left(\frac{1}{b-a+1}\right) \left[\text{Using (9.16)}\right]$$

$$= \frac{1}{b-a+1}\sum_{x=a}^{b} x \qquad \left[\begin{array}{l}\because \dfrac{1}{b-a+1} \text{ is independent of the} \\ \text{variable x of the summation}\end{array}\right]$$

$$= \frac{1}{b-a+1}\left[a + (a+1) + (a+2) + \ldots + b\right]$$

$$= \frac{1}{b-a+1}\left[\frac{b-a+1}{2}(a+b)\right] \qquad \left[\begin{array}{l}\because \text{Sum of n terms of an AP} \\ a + (a+d) + (a+2d) + \ldots + l = \dfrac{n}{2}(a+l) \\ a = \text{first term of AP},\ n = \text{number of terms} \\ \text{and } l = \text{last term of the given AP}\end{array}\right]$$

$$= \frac{a+b}{2}$$

$$\text{First raw moment} = \mu_1' = \frac{a+b}{2} \qquad \ldots (9.36)$$

We know that the first raw moment is mean of the distribution. Therefore,

mean or expected value of the discrete uniform distribution is given by

$$\mu_1' = E(X) = \frac{a+b}{2} \qquad \ldots (9.37)$$

$$\text{Second raw moment} = \mu_2' = E\left(X^2\right) = \sum_{x=a}^{b} x^2\,\mathcal{P}\left(X = x\right) = \sum_{x=a}^{b} x^2\left(\frac{1}{b-a+1}\right)\left[\begin{array}{l}\text{Using} \\ (9.16)\end{array}\right]$$

$$= \frac{1}{b-a+1}\sum_{x=a}^{b} x^2 \quad \left[\begin{array}{l}\because \dfrac{1}{b-a+1} \text{ is independent of the} \\ \text{variable x of the summation}\end{array}\right]$$

Putting x = y + a – 1

When $x = a \Rightarrow y = 1$, and when $x = b \Rightarrow y = b - a + 1 = N$ [Using (9.17)]

$$\therefore\ \mu_2' = \frac{1}{b-a+1}\sum_{y=1}^{N}(y+a-1)^2$$

$$= \frac{1}{N}\sum_{y=1}^{N}\left[y^2 + 2(a-1)y + (a-1)^2\right]^2 \quad [\because b-a+1 = N \text{ refer to } (9.17)]$$

$$= \frac{1}{N}\left[\sum_{y=1}^{N}y^2 + 2(a-1)\sum_{y=1}^{N}y + \sum_{y=1}^{N}(a-1)^2\right]$$

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

$$\left[\because 1+2+3+\ldots+n = \frac{n(n+1)}{2} \text{ and } 1^2+2^2+3^2+\ldots+n^2 = \frac{n(n+1)(2n+1)}{6}\right]$$

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

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

$$= (N+1)\left[\frac{2N+6a-5}{6}\right] + (a-1)^2$$

$$\Rightarrow\ \mu_2' = (N+1)\left[\frac{2N+6a-5}{6}\right] + (a-1)^2 \qquad \ldots (9.38)$$

Now, using (9.36) and (9.38) in (7.65) variance of random variable X is given by

$$\text{Variance of } X = \mu_2 = \mu_2' - (\mu_1')^2 = (N+1)\left[\frac{2N+6a-5}{6}\right] + (a-1)^2 - \left(\frac{a+b}{2}\right)^2$$

$$= (N+1)\left[\frac{2N+6a-5}{6}\right] + \left(a-1+\frac{a+b}{2}\right)\left(a-1-\frac{a+b}{2}\right) \quad \left[\begin{array}{l}\because a^2-b^2 \\ = (a-b)(a+b)\end{array}\right]$$

$$= (N+1)\left[\frac{2N+6a-5}{6}\right] + \left(\frac{2a-2+a+b}{2}\right)\left(\frac{2a-2-a-b}{2}\right)$$

$$= (N+1)\left[\frac{2N+6a-5}{6}\right] + \left(\frac{3a-2+b}{2}\right)\left(\frac{a-b-2}{2}\right)$$

$$= (N+1)\left[\frac{2N+6a-5}{6}\right] + \left(\frac{-3a+2-b}{2}\right)\left(\frac{b-a+2}{2}\right)$$

$$= (N+1)\left[\frac{2N+6a-5}{6}\right] + \left(\frac{-3a+2-b}{2}\right)\left(\frac{N+1}{2}\right) \quad \left[\begin{array}{l}\because b-a+1 = N \\ \Rightarrow b-a+2 = N+1\end{array}\right]$$

$$= (N+1)\left[\frac{2N+6a-5}{6} + \frac{-3a+2-b}{4}\right] = (N+1)\left[\frac{4N+12a-10-9a+6-3b}{12}\right]$$

$$= (N+1)\left[\frac{4N+3a-3b-4}{12}\right] = (N+1)\left[\frac{4N-3(b-a)-4}{12}\right]$$

$$= (N+1)\left[\frac{4N-3(N-1)-4}{12}\right] \quad [\because b-a+1 = N \Rightarrow b-a = N-1]$$

$$= \frac{(N+1)(N-1)}{12} = \frac{N^2-1}{12} = \frac{(b-a+1)^2-1}{12}$$

$\therefore$ Variance of discrete uniform distribution $X = \mu_2 = \dfrac{(b-a+1)^2 - 1}{12}$ … (9.39)

We know that standard deviation of X is positive square root of variance of X.

Hence, $SD(X) = \sqrt{\text{Variance of } X} = \sqrt{\dfrac{(b-a+1)^2 - 1}{12}}$ … (9.40)

$$SD(X) = \sigma = \sqrt{\frac{(b-a+1)^2 - 1}{12}}$$

Similarly, we can obtain other measures of our interest. But keeping applied nature of the programme in view we are not focusing on proof of each measure. Some commonly used summary measures of uniform distribution are shown in Table 9.4 given as follows.

Table 9.4: Summary measures of discrete uniform distribution

Name of measure	Formula	Name of measure	Formula
Mean	$\frac{a+b}{2}$	Standard deviation	$SD(X) = \sigma = \sqrt{\frac{(b-a+1)^2 - 1}{12}}$
Median	$\frac{a+b}{2}$	MGF	$\frac{e^{at}\left(1-e^{bt}\right)}{(b-a+1)\left(1-e^{t}\right)}$
Mode	Does not exist since all values have equal probabilities. It means all values are mode. It is not providing the kind of information mode is known for. So, in such situations some authors say that all values are mode on the other hand some authors say that mode does not exist.	Skewness	0
Variance	$\mu_2 = \frac{(b-a+1)^2 - 1}{12}$	Kurtosis	$\frac{6}{5}\left[\frac{(b-a+1)^2 + 1}{(b-a+1)^2 - 1}\right]$

9.4 STORY, DEFINITION, PMF AND CDF OF BERNOULLI DISTRIBUTION

To understand Bernoulli distribution first, we have to understand when a trial in a random experiment is said to be a Bernoulli trial. Performing a random experiment is known as a trial. For example, tossing a coin is a trial. If each trial of a random experiment is termed in one of the two possible categories traditionally known as a success or a failure then such a trial is known as **Bernoulli trial**. If we call getting a head as success and getting a tail as a failure then tossing a coin is a Bernoulli trial because either outcome will be head (success) or tail (failure). There are two important points that should be noted here are:

- If a random experiment has more than two outcomes in each trial then such trials can also be considered as Bernoulli trials provided, we have to define success and failure events E and F respectively such that each trial either termed as success or failure. That is $E \cap F = \phi$ and $E \cup F = \Omega$. For example, in the random experiment of throwing a die there are six possible outcomes 1 or 2 or 3 or 4 or 5 or 6. If we are interested in getting a multiple of 3 then event E will be getting 3 or 6 and event F not multiple of 3 will be getting 1 or 2 or 4 or 5. Thus, E = {3, 6} and F = {1, 2, 4, 5} and probability of success $= \mathcal{P}(E) = \frac{2}{6} = \frac{1}{3}$, while probability of failure $= \mathcal{P}(F) = \frac{4}{6} = \frac{2}{3}$.

Generally probability of success is denoted by p and probability of failure is denoted by 1 – p or sometimes by q. So, in this example,

$$p = \frac{1}{3} \text{ and } 1 - p = \frac{2}{3}. \qquad \ldots (9.41)$$

- Another point that should be clear to you at this point of time is here success does not mean the kind of success, we are used to in real life. For example, in real life success means producing a good product/item but in probability theory, we can also define success as getting a defective item or good item depends on our event of interest. Success can also be defined as detecting cancer or getting an accident, etc. In real life, we do not consider happening of such event as success. So, keep this distinction of success in probability and real life in your mind. **In probability success totally depends on our event of interest it may or may not contradict the meaning of success in real life**. … (9.42)

Now, we can write the story of Bernoulli distribution as follows.

Story of Bernoulli Distribution: If we perform a random experiment and the realisation of a trial has only two categories success or failure, then probability distribution of a random variable which takes value 1 if outcome is a success and 0 if outcome is a failure is known as Bernoulli distribution. … (9.43)

So, by definition of a discrete probability distribution, we mean the specification of Ω and $\mathcal{P}$.

Now, we define Bernoulli distribution as follows.

Definition and PMF of Bernoulli Distribution: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \{\text{success, failure}\}$ contains only two types of outcomes traditionally known as success and failure. Let X be a random variable defined on the sample space Ω by $X(\text{success}) = 1$ and $X(\text{failure}) = 0$. So, random variable X assumes only two values 0 and 1. We say that the random variable X follows Bernoulli distribution if probability measure $\mathcal{P}$ assigned probabilities p and 1 – p to success and failure respectively, i.e., if

$$\mathcal{P}(X = x) = \begin{cases} p, & \text{if } x = 1 \\ 1 - p, & \text{if } x = 0 \end{cases} \qquad \ldots (9.44)$$

where $0 \le p \le 1$ … (9.45)

So, PMF of Bernoulli random variable X is given by

$$\mathcal{P}(X = x) = \begin{cases} p^x (1-p)^{1-x}, & \text{if } x = 0, 1 \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9.46)$$

If random variable X follows Bernoulli distribution with probability of success p, then p is known as parameter of Bernoulli distribution and is denoted by writing X ~ Bern(p) or X ~ Bernoulli(p). In this course, we will use the notation X ~ Bern(p). … (9.47)

Like discrete uniform distribution case, we read Bern(p) as X is distributed as Bernoulli distribution with parameter p. Or we read it as X follows Bernoulli distribution with parameter p. … (9.48)

Since the statistical software used for hands on practice for the lab courses of this programme is R programming language, so you should also know its notation in R. In R notation that is used for Bernoulli distribution is bern(prob) in the package Rlab, where prob represents value of p. In fact, like any probability distribution there are four functions for Bernoulli distribution namely dbern(x, min, max, …), pbern(q, min, max, …), qbern(p, min, max, …) and rbern(n, min, max, …). We have already explained meaning of these functions in (9.22) to (9.25). ... (9.49)

Let us check the **validity of the PMF of the Bernoulli distribution**.

We have to check two conditions non-negativity and normality refer to (4.62).

(1) **Non-negativity**: Since $0 \le p \le 1$, so $0 \le 1-p \le 1$. … (9.50)

(2) **Normality**: Since $\mathcal{P}(X=1)=p$ and $\mathcal{P}(X=0)=1-p$.

But $p+(1-p)=1$. … (9.51)

This proves that sum of all probabilities of Bernoulli distribution is 1.

Hence, we can say that PMF given by (9.46) is a valid PMF.

Now, we define CDF of Bernoulli distribution.

CDF of Bernoulli Distribution: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \{\text{success, failure}\}$ contains only two types of outcomes traditionally known as success and failure. Let X be a random variable defined on the sample space Ω by $X(\text{success})=1$ and $X(\text{failure})=0.$ So, random variable X assumes only two values 0 and 1. The random variable X follows Bernoulli distribution if its PMF is given by

$$\mathcal{P}(X=x)=\begin{cases} p^x(1-p)^{1-x}, & \text{if } x=0,1 \\ 0, & \text{otherwise} \end{cases}$$

Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x)=\mathcal{P}(X \le x)=\begin{cases} 0, & \text{if } x<0 \\ 1-p, & \text{if } 0 \le x < 1 \\ 1, & \text{if } x \ge 1 \end{cases} \quad \text{… (9.52)}$$

Let us do one example.

Example 2: Plot PMF of Bernoulli random variable when p = 0, 0.25, 0.5, 0.75, 0.9, and 1.

Solution: PMF of Bernoulli random variable when p = 0, 0.25, 0.5, 0.75, 0.9 and 1 are shown in Fig. 9.3 (a), (b), (c), (d), (e), and (f) respectively as follows.

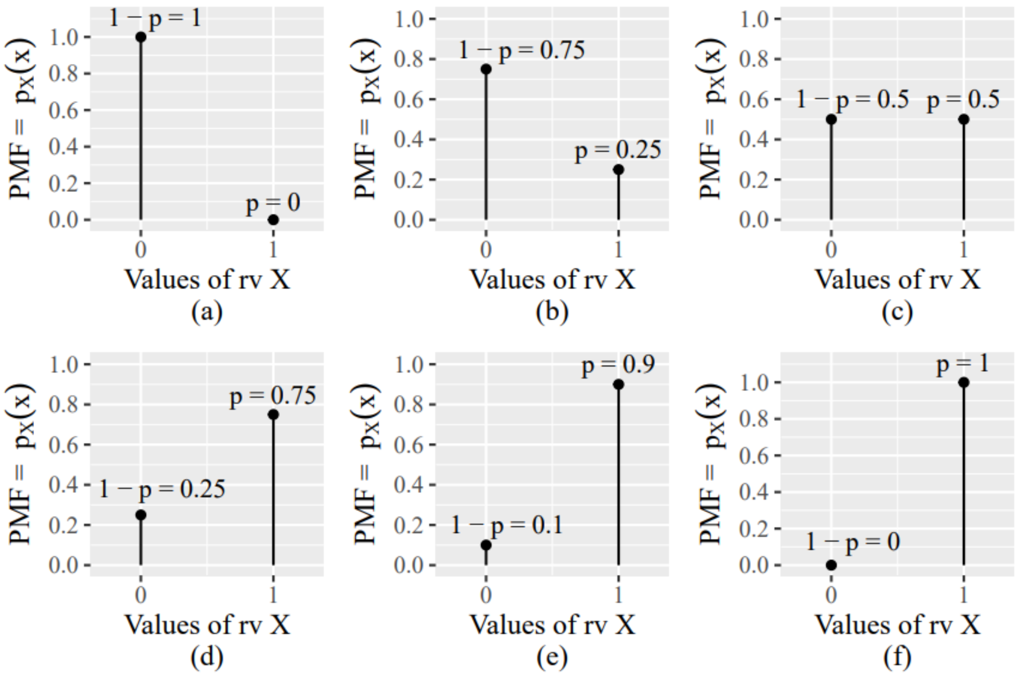


Fig. 9.3: Visualisation of PMF of Bernoulli distribution when p is (a) 0 (b) 0.25 (c) 0.5 (d) 0.75 (e) 0.9 (f) 1

Example 3: Plot PMF and CDF of Bernoulli random variable when p = 0.6.

Solution: PMF and CDF of Bernoulli random variable when p = 0.6 are shown in Fig. 9.4 (a) and (b) respectively as follows.

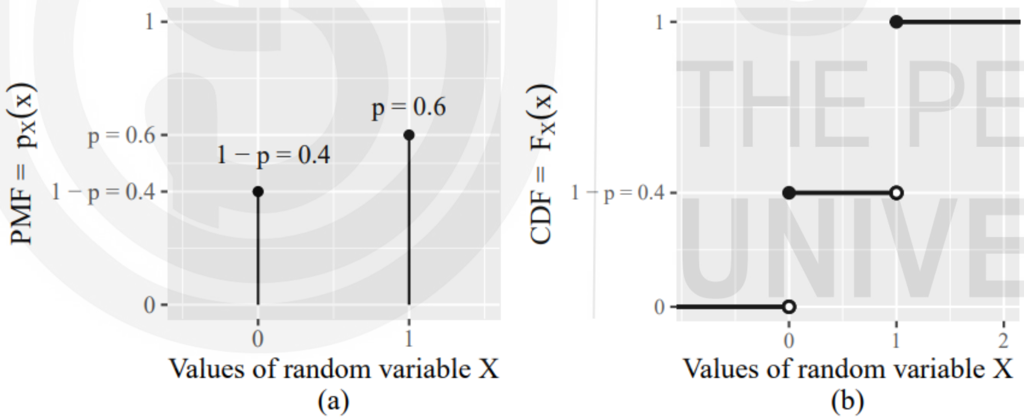


Fig. 9.4: Visualisation of (a) PMF (b) CDF of Bernoulli rv X when p = 0.6

Remark 2: In the special case p = 0 we say that X is a constant random variable having only one value 0 with probability 1 refer to Fig. 9.3 (a). Similarly, in the other special case p = 1 we say that X is a constant random variable having only one value 1 with probability 1 refer to Fig. 9.3 (f).… (9.53)

9.5 MGF AND OTHER SUMMARY MEASURES OF BERNOULLI DISTRIBUTION

In the previous section, you have studied PMF and CDF of Bernoulli distribution. In this section, we want to obtain MGF and some other summary measure of Bernoulli distribution like mean, median, variance, etc. Let us first obtain MGF of Bernoulli distribution.

Calculation of MGF

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{1} e^{tx}\, p_X(x) = \sum_{x=0}^{1} e^{tx}\, p^x (1-p)^{1-x}$$

$$= e^{t(0)}\, p^0 (1-p)^{1-0} + e^{t(1)}\, p^1 (1-p)^{1-1} = 1-p+p\,e^t$$

$$\Rightarrow M_X(t) = 1-p+p\,e^t \qquad \ldots (9.54)$$

$$\text{or } M_X(t) = q+p\,e^t \qquad \ldots (9.55)$$

where $q = 1-p$

Calculation of Some Summary Measures

Recall (7.57) to (7.60), all raw moments can be obtained by substituting t = 0 in the expressions of different derivatives of MGF. But we are interested only in the first four raw moments. So, let us differentiate (9.54) or (9.55) with respect to t successively four times, we get

$$M_X^{(1)}(t) = p\,e^t \quad \left[\because \text{ q is constant so } \frac{d}{dt}(q) = 0 \text{ and } \frac{d}{dt}\left(e^t\right) = e^t\right] \qquad \ldots (9.56)$$

$$M_X^{(2)}(t) = p\,e^t \qquad \ldots (9.57)$$

$$M_X^{(3)}(t) = p\,e^t \qquad \ldots (9.58)$$

$$M_X^{(4)}(t) = p\,e^t \qquad \ldots (9.59)$$

Now, $\mu_1' = M_X^{(1)}(0) = pe^0 = p.$ … (9.60)

Similarly, $\mu_2' = M_X^{(2)}(0) = p,\ \ \mu_3' = M_X^{(3)}(0) = p,\ \ \mu_4' = M_X^{(4)}(0) = p.$

Hence,

$$\left.\begin{array}{l}\textbf{Expected value of Bern(p)} = \textbf{Mean} = \mu_1' = p \\ \textbf{Variance of Bern(p)} = \mu_2 = \mu_2' - (\mu_1')^2 = p - p^2 = p(1-p)\end{array}\right\} \qquad \ldots (9.61)$$

We know that standard deviation of X is positive square root of variance of X. Hence, Standard Deviation $= SD(X) = \sqrt{\text{Variance of } X} = \sqrt{p(1-p)}$ … (9.62)

Before obtaining skewness and kurtosis first we have to obtain **third and fourth central moments** using (7.109) and (7.110) as follows.

$$\mu_3 = \mu_3' - 3\mu_2'\,\mu_1' + 2\left(\mu_1'\right)^3 = p - 3p^2 + 2p^3 = p\left(1-3p+2p^2\right) = p\left(2p^2 - 3p + 1\right)$$

$$= p\left[2p^2 - 2p - p + 1\right] = p\left[2p(p-1) - 1(p-1)\right] = p(p-1)(2p-1)$$

$$= p(1-p)(1-2p) \qquad \ldots (9.63)$$

$$\mu_4 = \mu_4' - 4\mu_3'\,\mu_1' + 6\mu_2'\,(\mu_1')^2 - 3(\mu_1')^4 = p - 4p^2 + 6p^3 - 3p^4 = p(1-4p+6p^2-3p^3)$$

Since p = 1 satisfies $1-4p+6p^2-3p^3$ so p – 1 is a factor of it. So, dividing using long division or synthetic division, we have

$$\mu_4 = p(p-1)\left(-3p^2+3p-1\right) = p(1-p)\left(1-3p+3p^2\right) \qquad \ldots (9.64)$$

$$\therefore \textbf{ Skewness } = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{p(1-p)(1-2p)}{[p(1-p)]^{3/2}} = \frac{1-2p}{\sqrt{p(1-p)}} \qquad \ldots (9.65)$$

and **Kurtosis** $= \frac{\mu_4}{(\mu_2)^2} = \frac{p(1-p)(1-3p+3p^2)}{[p(1-p)]^2} = \frac{1-3p+3p^2}{p(1-p)}$ … (9.66)

Calculation for Mode

From Fig, 9.3 (a) and (b) note that when p < 1/2 then $\mathcal{P}(X=0) > \mathcal{P}(X=1)$ so in this case mode will be 0.

From Fig, 9.3 (d), (e) and (f) note that when p > 1/2 then $\mathcal{P}(X=1) > \mathcal{P}(X=0)$ so in this case mode will be 1.

However, from Fig, 9.3 (c) note that when p = 1/2 then $\mathcal{P}(X=0) = \mathcal{P}(X=1)$ so in this case either we can say mode does not exist like discrete uniform distribution case or we can say that both 0 and 1 are mode. So, we have

$$\text{Mode of } X = \begin{cases} 0, & \text{if } p < 1/2 \\ 0, 1, & \text{if } p = 1/2 \\ 1, & \text{if } p > 1/2 \end{cases} \quad \dots (9.67)$$

Calculation for Median

We know that if M is median of a probability distribution of a random variable X then we have

$$\mathcal{P}(X \le M) \ge \frac{1}{2} \text{ and } \mathcal{P}(X \ge M) \ge \frac{1}{2} \quad \dots (9.68)$$

From Fig, 9.3 (a) and (b) note that when p < 1/2 then

$$\mathcal{P}(X \le 0) = \mathcal{P}(X=0) = 1-p \ge \frac{1}{2} \quad \left[\because p < \frac{1}{2}\right] \text{ and}$$

$$\mathcal{P}(X \ge 0) = \mathcal{P}(X=0) + \mathcal{P}(X=1) = 1-p+p = 1 \ge \frac{1}{2}$$

So, in this case median is 0. Similarly, when p > 1/2, then median will be 1 and when p = 1/2, then any value between 0 and 1 including 0 and 1 will satisfy (9.68). Hence, median of Bernoulli random variable is given by

$$\text{Median of } X = \begin{cases} 0, & \text{if } p < 1/2 \\ [0, 1], & \text{if } p = 1/2 \\ 1, & \text{if } p > 1/2 \end{cases} \quad \dots (9.69)$$

Let us put these calculated summary measures of Bernoulli distribution in Table 9.5 given as follows.

Table 9.5: Summary measures of Bernoulli distribution

Name of measure	Formula	Name of measure	Formula
Mean	p	Standard deviation	$SD(X) = \sigma = \sqrt{p(1-p)}$
Median	$\begin{cases} 0, & \text{if } p < 1/2 \\ [0, 1], & \text{if } p = 1/2 \\ 1, & \text{if } p > 1/2 \end{cases}$	MGF	$1-p+pe^t$ or $q+pe^t$ where $q = 1-p$

Mode	$\begin{cases} 0, & \text{if } p < 1/2 \\ 0, 1, & \text{if } p = 1/2 \\ 1, & \text{if } p > 1/2 \end{cases}$	Skewness	$\frac{1-2p}{\sqrt{p(1-p)}}$
Variance	$\mu_2 = p(1-p)$	Kurtosis	$\frac{1-3p+3p^2}{p(1-p)}$

9.6 APPLICATIONS AND ANALYSIS OF DISCRETE UNIFORM AND BERNOULLI DISTRIBUTIONS

In this section, we will apply discrete uniform and Bernoulli distributions to solve some problems where assumptions of these distribution make sense. We will also do some analysis of these two distributions.

Example 4: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and E be an event then find the probability distribution of the indicator random variable $I_E : \Omega \to \{0, 1\}$ of event E.

Solution: In Sec. 4.7 of Unit 4, we have discussed indicator random variables and their properties. From the discussion of Sec. 4.7, we know that indicator random variable $I_E : \Omega \to \{0, 1\}$ of event E is defined by

$$I_E(\omega) = \begin{cases} 1, & \text{if } \omega \in E \\ 0, & \text{if } \omega \notin E \end{cases} \quad \omega \in \Omega \qquad \dots (9.70)$$

If we denote probability of event E by p, then we have

$$\mathcal{P}(E) = p \text{ and } \mathcal{P}(E^c) = 1 - p. \qquad \dots (9.71)$$

Using (9.71) in (9.70), PMF of the indicator random variable I_E is given by

$$p_{I_E}(x) = \mathcal{P}(I_E = x) = \begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0 \end{cases} \qquad \dots (9.72)$$

which is a PMF of a Bernoulli distribution with parameter $p = \mathcal{P}(E)$. Hence,

$$I_E \sim \text{Bern}(p = \mathcal{P}(E)). \qquad \dots (9.73)$$

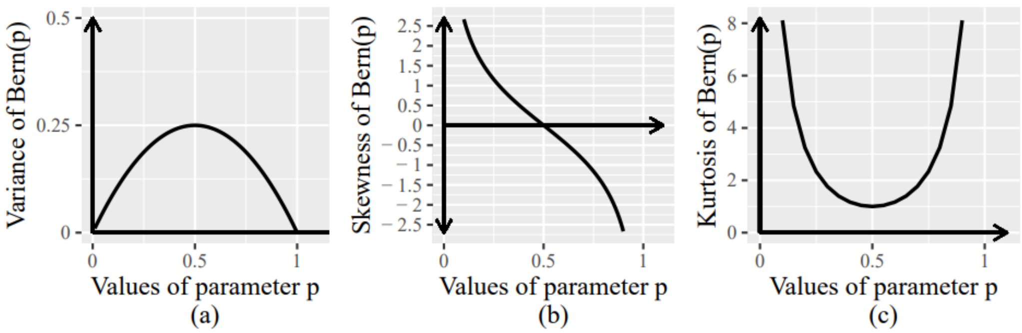


Fig. 9.5: Visualisation of (a) variance (b) skewness (c) kurtosis with respect to values of p of Bernoulli distribution with parameter p

Example 5: **Analysis of some measures of Bern(p)**: If X ~ Bern(p), then discuss the behaviour of variance, skewness and kurtosis graphically with respect to values of p. Also, explain main points of graphical behaviour.

Solution: Graphs of variance, skewness and kurtosis with respect to values of p are shown in Fig. 9.5 (a), (b) and (c) respectively.

Comment on Observing Graphical Behaviour of Variance: We know that variance measures the spread of the distribution. From Fig. 9.5 (a), we observe the following facts.

- Note that if p = 0 then variance = 0. It should be because when p = 0 then Bernoulli random variable X will take value 0 only with probability 1 (= 1 – p = 1 – 0). So, there is no variation in the values of X and hence variance should be 0. … (9.74)

- If p = 1 then variance is also 0. It should be because when p = 1 then Bernoulli random variable X will take value 1 only with probability 1. So, there is no variation in the values of X and hence variance should be 0.

  … (9.75)

- Maximum value of variance is 0.25 and it is obtained when p = 0.5. Obviously, when p = 0.5 then two values 0 and 1 of X will have equal probability 0.5 of occurrence and hence variation in the values of Bernoulli random variable X will be maximum. … (9.76)

- Finally, variance of Bernoulli random variable decreases as p decreases from 0.5 to 0 or p increases from 0.5 to 1. … (9.77)

Comment on Observing Graphical Behaviour of Skewness: We know that skewness is a measure of degree of asymmetry in the distribution of a random variable. We also know that:

(a) If skewness = 0 then distribution is perfectly symmetric about it mean.

… (9.78)

(b) If $-0.5 \leq \text{Skewness} \leq 0.5,$ then distribution is approximately symmetric.

… (9.79)

(c) If $-1 \leq \text{Skewness} \leq -0.5$ or $0.5 \leq \text{Skewness} \leq 1,$ then distribution is moderately skewed. … (9.80)

(d) If $\text{Skewness} < -1$ or $\text{Skewness} > 1,$ then distribution is highly skewed.

… (9.81)

Keeping this in view from Fig. 9.5 (b), we observe the following facts.

- Note that skewness is 0 when p = 0.5. It implies PMF of X should be symmetrical about its mean = p = 0.5 which is true refer to Fig. 9.3 (c).

  … (9.82)

- As $p \to 0$ then skewness tends to $\infty.$ It means distribution of Bernoulli random variable X will be positively skewed refer to Fig. 9.3 (a) and (b).

  … (9.83)

- As $p \to 1$ then skewness tends to $-\infty$. It means distribution of Bernoulli random variable X will be negatively skewed which is true refer to Fig. 9.3 (d), (e) and (f). … (9.84)

- So, Bernoulli distribution will be: approximately symmetric if

$$\frac{17-\sqrt{17}}{34} \le p \le \frac{17+\sqrt{17}}{34} \left[\begin{array}{l} \because \dfrac{1-2p}{\sqrt{p(1-p)}} = \dfrac{1}{2} \Rightarrow 1+4p^2-4p = \dfrac{p-p^2}{4} \\ \Rightarrow 17p^2-17p+4=0 \Rightarrow p = \dfrac{-b \pm \sqrt{b^2-4ac}}{2a} = \dfrac{17 \pm \sqrt{17}}{34} \end{array}\right]$$

… (9.85)

moderately skewed if

$$\begin{array}{l} \dfrac{5-\sqrt{5}}{10} \le p \le \dfrac{17-\sqrt{17}}{34} \\ \text{Or } \dfrac{17+\sqrt{17}}{34} \le p \le \dfrac{5+\sqrt{5}}{10} \end{array} \left[\begin{array}{l} \because \dfrac{1-2p}{\sqrt{p(1-p)}} = 1 \Rightarrow 1+4p^2-4p = p-p^2 \\ \Rightarrow 5p^2-5p+1=0 \Rightarrow p = \dfrac{-b \pm \sqrt{b^2-4ac}}{2a} = \dfrac{5 \pm \sqrt{5}}{10} \end{array}\right]$$

If we want to write in decimal form then

$0.2763932 \le p \le 0.3787322$ and $0.6212678 \le p \le 0.7236068$ … (9.86)

highly skewed if $p < 0.2763932$ or $p > 0.7236068$. … (9.87)

Comment on Observing Graphical Behaviour of Kurtosis: Till 2014 it was stated that kurtosis is a measure of both peak of the distribution as well as fatness of the tails of the distribution. But in 2014, Peter H. Westfall published a paper in the journal of The American Statistician (refer last page of this unit for reference of this paper) where he showed that kurtosis only measures fatness of the tails. Value of kurtosis of a distribution vary from 1 to ∞. We also know that value of kurtosis for normal distribution is 3. So, classification of kurtosis is done as follows. … (9.88)

(a) If a distribution like normal has value of kurtosis = 3 then we call it mesokurtic. … (9.89)

(b) If a distribution has value of kurtosis > 3 then we call it leptokurtic. It means tails of this distribution are fat than normal distribution. In other words, probability under the tails of this distribution is more than the tails of normal distribution. … (9.90)

(c) If a distribution has value of kurtosis < 3 then we call it platykurtic. It means tails of this distribution are less fat than normal distribution. In other words, probability under the tails of this distribution is less than the tails of normal distribution. … (9.91)

Keeping this in view from Fig. 9.5 (c) we observe the following facts.

- Distribution of Bern(p) will be mesokurtic if

$$p = \frac{3 \pm \sqrt{3}}{6} \left[\begin{array}{l} \because \dfrac{1-3p+3p^2}{p(1-p)} = 3 \Rightarrow 1-3p+3p^2 = 3p-3p^2 \\ \Rightarrow 6p^2-6p+1=0 \Rightarrow p = \dfrac{-b \pm \sqrt{b^2-4ac}}{2a} = \dfrac{3 \pm \sqrt{3}}{6} \end{array}\right] \quad \dots (9.92)$$

If we want to write value of p in decimal form using R as a calculator, we have p = 0.2113249 and p = 0.7886751.

- Similarly, distribution of Bern(p) will be leptokurtic if

$$p > \frac{3+\sqrt{3}}{6} \text{ or } p < \frac{3-\sqrt{3}}{6} \qquad \ldots (9.93)$$

- and distribution of Bern(p) will be platykurtic if value of p satisfies

$$\frac{3-\sqrt{3}}{6} < p < \frac{3+\sqrt{3}}{6} \qquad \ldots (9.94)$$

Example 6: In a sector of a city 1000 families are living. It is known that 1 out of 60 parents have a twin. Out of these 1000 families there are 900 families each have parents of exactly two generations like grandfather/grandmother his/her children and his/her grandchildren. There are 10 families each having parents of exactly three generations. Remaining 90 families have parents of only one generation. Assume that all parents are equally likely to be selected. A parent is selected at random. (a) Find the expected value that a selected parent has a twin. (b) Find expected number of parents having twin.

Solution: According to the problem total number of parents (N) in these 1000 families is given by

$$N = 900 \times 2 + 10 \times 3 + 90 \times 1 = 1800 + 30 + 90 = 1920 \qquad \ldots (9.95)$$

Let $X_k,\ k = 1, 2, 3, 4, \ldots, 1920$ be the indicator random variable whether k^{th} parent has twin or not. So, using (9.73), we have

$$X_k \sim \text{Bern}\left(p = \frac{1}{60}\right),\ k = 1, 2, 3, 4, \ldots, 1920 \qquad \ldots (9.96)$$

(a) Now, using (9.96), expected value that a selected parent has twin is given by

$$E(X_k) = p = \frac{1}{60},\ k = 1, 2, 3, 4, \ldots, 1920 \qquad \ldots (9.97)$$

(b) Finally, expected number of parents having twin is given by

$$E\left(\sum_{k=1}^{1920} X_k\right) = \sum_{k=1}^{1920} E(X_k) \qquad \left[\begin{array}{l}\text{Using addition theorem of}\\ \text{expectation refer to (7.30)}\end{array}\right]$$

$$= \sum_{k=1}^{1920}\left(\frac{1}{60}\right) \qquad [\text{Using (9.97)}]$$

$$= 1920 \times \frac{1}{60} \qquad \left[\because \sum_{k=1}^{n} a = na, \text{ if } a \text{ is independent of } k\right]$$

$$= 32 \qquad \ldots (9.98)$$

Hence, expected number of parents who have twin is 32 out of 1920 parents of 1000 families who are living in this particular sector.

Example 7: Suppose you are interested in the distribution formed by the last digits of all the mobile numbers of a particular company. Without collecting data what distribution the last digit of mobile number may supposed to follow. Write PMF of this distribution.

Solution: Let X denote the last digit of the mobile numbers of the company in which we are interested. So, X can take values 0, 1, 2, 3, 4, …, 9. Without collecting data we are expecting that all the digits 0 to 9 are equally likely to be the last digit of a mobile number. So, suitable probability distribution for X is discrete uniform with parameters a = 0 and b = 9. So, PMF of the random variable X is given by

$$p_X(x) = \mathcal{P}\left(X = x\right) = \begin{cases} \dfrac{1}{b-a+1}, & \text{if } x = a, a+1, a+2, \ldots, b \\ 0, & \text{otherwise} \end{cases}$$

In our case a = 0, b = 9, so PMF of X is given by

$$p_X(x) = \mathcal{P}\left(X = x\right) = \begin{cases} \dfrac{1}{9-0+1}, & \text{if } x = 0, 1, 2, \ldots, 9 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Or } p_X(x) = \mathcal{P}\left(X = x\right) = \begin{cases} \dfrac{1}{10}, & \text{if } x = 0, 1, 2, \ldots, 9 \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9.99)$$

Example 8: Suppose we have n numbers 1, 2, 3, 4, …, n. Let S_n denote the set of all permutations of these n numbers taken all at a time. That is

$$S_n = \left\{x_1 x_2 x_3 x_4 \ldots x_n : x_i \in \{1, 2, 3, 4, \ldots, n\}, 1 \le i \le n, x_i \ne x_j \text{ if } i \ne j\right\} \qquad \ldots (9.100)$$

If X_i denotes the value of x_i, $1 \le i \le n$ in a randomly selected permutation from the set S_n, then find the probability distribution of X_i.

Solution: From school mathematics we know that total number of permutations of n things taken all at a time is $\underline{|n}$. For example, all the possible permutations of three number 1, 2, 3 taken all at a time are: 123, 132, 213, 231, 312 and 321 which are $6 = \underline{|3}$ in numbers. So, in this case $S_3 = \left\{123, 132, 213, 231, 312, 321\right\}$.

Suppose randomly selected permutation from S_3 is 312, then $X_1 = x_1 = 3,\ X_2 = x_2 = 1,\ X_3 = x_3 = 2.$ Similarly, if randomly selected permutation from S_3 is 213, then $X_1 = x_1 = 2,\ X_2 = x_2 = 1,\ X_3 = x_3 = 3.$

Now, let us discuss general case. Let x be any but a fixed number from 1, 2, 3, …, n then

$$\mathcal{P}\left(X_i = x\right) = \frac{\underline{|n-1}}{\underline{|n}},\ x = 1, 2, 3, \ldots, n \quad \left[\begin{array}{l} \because n\left(S_n\right) = \underline{|n}. \text{ Now if } x \text{ is a fix number} \\ \text{among } 1, 2, 3, \ldots, n, \text{ then remaining } n-1 \\ \text{positions can be filled up with remaining} \\ n-1 \text{ numbers.} \\ \text{This can be done in } \underline{|n-1} \text{ ways} \end{array}\right]$$

$$= \frac{\lfloor n-1}{n\lfloor n-1} = \frac{1}{n}$$

Hence, PMF of X_i is given by

$$p_{X_i}(x) = \mathcal{P}\left(X_i = x\right) = \begin{cases} \frac{1}{n}, & \text{if } x = 1, 2, 3, 4, \ldots, n \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (9.101)$$

which is PMF of discrete uniform distribution with parameters 1 and n, Hence, each X_i follows dunif(1, n).

Now, you can try the following two Self-Assessment Questions.

SAQ 1

Find expected value and standard deviation of the random variable discussed in Example 7. Also, give interpretation of standard deviation.

SAQ 2

Find skewness and kurtosis of the random variable discussed in Example 7. Also, give interpretation of the kurtosis.

9.7 SUMMARY

A brief summary of what we have covered in this unit is given as follows:

- **Definition of Discrete Uniform Distribution**: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \{a, a+1, a+2, a+3, \ldots, b-1, b\}$ contains $b - a + 1 = N$ (say), number of outcomes of a random experiment which are finite in number. Let X be a random variable defined on the sample space Ω by $X(\omega) = \omega \ \forall \ \omega \in \Omega$. We say that the random variable X follows discrete uniform distribution if probability measure $\mathcal{P}$ assigned equal probability to each value of X, i.e., if

$$\mathcal{P}\left(X = x\right) = \frac{1}{b-a+1}, \ x = a, a+1, a+2, a+3, \ldots, b \qquad \text{where } b - a + 1 = N$$

- **PMF** of discrete uniform random variable X is given by

$$\mathcal{P}\left(X = x\right) = \begin{cases} \frac{1}{b-a+1}, & x = x = a, a+1, a+2, a+3, \ldots, b \\ 0, & \text{otherwise} \end{cases}$$

- **CDF of Discrete Uniform Distribution**: Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}\left(X \le x\right) = \sum_{k=0}^{[x]} \mathcal{P}\left(X = k\right) = \begin{cases} 0, & \text{if } x < a \\ \frac{[x] - a + 1}{b-a+1}, & \text{if } a \le x < b \\ 1, & \text{if } x \ge b \end{cases}$$

- If we have a = b then in this special case we say that X is a constant random variable having only one value ‘a’ with $\mathcal{P}(X = a) = 1$.

- **Summary measures of discrete uniform distribution**

Name of measure	Formula	Name of measure	Formula
Mean	$\frac{a+b}{2}$	Standard deviation	$SD(X) = \sigma = \sqrt{\frac{(b-a+1)^2 - 1}{12}}$
Median	$\frac{a+b}{2}$	MGF	$\frac{e^{at}\left(1-e^{bt}\right)}{(b-a+1)\left(1-e^{t}\right)}$
Mode	Does not exist since all values have equal probabilities. It means all values are mode. It is not providing the kind of information mode is known for. So, in such situations some authors say that all values are mode on the other hand some authors say that mode does not exist.	Skewness	0
Variance	$\mu_2 = \frac{(b-a+1)^2 - 1}{12}$	Kurtosis	$\frac{6}{5}\left[\frac{(b-a+1)^2 + 1}{(b-a+1)^2 - 1}\right]$

- If each trial of the random experiment is termed in one of the two possible categories traditionally known as a success or a failure then such a trial is known as **Bernoulli trial**.

- **Definition of Bernoulli Distribution**: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \{\text{success, failure}\}$ contains only two types of outcomes traditionally known as success and failure. Let X be a random variable defined on the sample space Ω by $X(\text{success}) = 1$ and $X(\text{failure}) = 0$. So, random variable X assumes only two values 0 and 1. We say that the random variable X follows Bernoulli distribution if probability measure $\mathcal{P}$ assigned probabilities p and 1 – p to success and failure respectively, i.e., if

- $\mathcal{P}(X = x) = \begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0 \end{cases}$ where $0 \le p \le 1$

- **PMF** of Bernoulli random variable X is given by

$$\mathcal{P}(X = x) = \begin{cases} p^x(1-p)^{1-x}, & \text{if } x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

- **CDF of Bernoulli Distribution**: Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}(X \le x) = \begin{cases} 0, & \text{if } x < 0 \\ 1-p, & \text{if } 0 \le x < 1 \\ 1, & \text{if } x \ge 1 \end{cases}$$

- **Summary measures of Bernoulli distribution**

Name of measure	Formula	Name of measure	Formula
Mean	p	Standard deviation	$SD(X) = \sigma = \sqrt{p(1-p)}$
Median	$\begin{cases} 0, & \text{if } p < 1/2 \\ [0, 1], & \text{if } p = 1/2 \\ 1, & \text{if } p > 1/2 \end{cases}$	MGF	$1-p+pe^t$ or $q+pe^t$ where $q = 1-p$
Mode	$\begin{cases} 0, & \text{if } p < 1/2 \\ 0, 1, & \text{if } p = 1/2 \\ 1, & \text{if } p > 1/2 \end{cases}$	Skewness	$\dfrac{1-2p}{\sqrt{p(1-p)}}$
Variance	$\mu_2 = p(1-p)$	Kurtosis	$\dfrac{1-3p+3p^2}{p(1-p)}$

9.8 TERMINAL QUESTIONS

1. A delivery company deliver packed food items to the customers. Past experience shows that company got 0, 1, 2, 3 and 4 delivery orders between 9 am to 10 pm all are equally likely. Find the expected number of delivery orders that this company may get between 9 am to 10 am in a randomly selected day. Also find probability that the company gets at least as many orders as the expected value.

2. If $X \sim \text{dunif}(a, b)$ and E(X) = 11, $SD(X) = \sigma = \sqrt{2}$ then find parameters of the distribution of X.

3. If $X \sim \text{Bern}(p)$ and $SD(X) = \sigma = 0.25$ then find parameter of the Bernoulli distribution of X.

9.9 SOLUTIONS/ANSWERS

Self-Assessment Questions (SAQs)

1. We know that expected value and variance of a discrete uniform distribution with parameters a and b are given by

Expected value of random variable X $= \text{Mean} = \dfrac{a+b}{2}$ and

$$SD(X) = \sigma = \sqrt{\frac{(b-a+1)^2 - 1}{12}}$$

In our case a = 0, b = 9, so, we have

Expected value of random variable X $= \text{Mean} = \dfrac{0+9}{2} = \dfrac{9}{2} = 4.5.$

Similarly, $SD(X) = \sigma = \sqrt{\dfrac{(9-0+1)^2-1}{12}} = \sqrt{\dfrac{99}{12}} = 2.872281.$

Interpretation of standard deviation is that on average values of X are 2.872281 units away from the expected value 4.5.

2. We know that skewness of a discrete uniform distribution is always 0. It tells us that PMF of discrete uniform random variable is symmetric about its mean.

We know that kurtosis of a discrete uniform distribution with parameters a and b is given by

$$\text{Kurtosis} = \frac{6}{5}\left[\frac{(b-a+1)^2+1}{(b-a+1)^2-1}\right]$$

In our case a = 0, b = 9, so, we have

$$\text{Kurtosis} = \frac{6}{5}\left[\frac{(9-0+1)^2+1}{(9-0+1)^2-1}\right] = \frac{6}{5}\left(\frac{101}{99}\right) = 1.224242.$$

Interpretation: Since value of kurtosis is < 3 so its distribution will be platykurtic.

Terminal Questions

1. The crucial points to identify which probability distribution is suitable for this problem are: (a) Number of outcomes are finite (b) All the outcomes are equally likely and (c) outcomes are independent.

Keeping this in view, we see that all the requirements of discrete uniform distribution are satisfied and hence it is a perfect situation where we can apply discrete uniform distribution with parameters a = 0 and b = 4. If X denotes the number of delivery orders that this company gets between 9 am and 10 am then X is a random variable and as explained it follows discrete uniform distribution. So, PMF of the random variable X is given by

$$p_X(x) = \mathcal{P}(X = x) = \begin{cases} \dfrac{1}{b-a+1}, & x = a, a+1, a+2, a+3, \ldots, b \\ 0, & \text{otherwise} \end{cases}$$

In our case a = 0 and b = 4, so

$$p_X(x) = \mathcal{P}(X = x) = \begin{cases} \dfrac{1}{4-0+1}, & x = 0, 1, 2, 3, 4 \\ 0, & \text{otherwise} \end{cases}$$

Or $p_X(x) = \mathcal{P}(X = x) = \begin{cases} \dfrac{1}{5}, & x = 0, 1, 2, 3, 4 \\ 0, & \text{otherwise} \end{cases}$ … (9.102)

We know that expected value of the discrete uniform distribution is given by $\dfrac{a+b}{2} = \dfrac{0+4}{2} = \dfrac{4}{2} = 2.$

Hence, required probability is given by

$$\mathcal{P}(X \geq 2) = \sum_{x=2}^{4} \mathcal{P}(X = x) = \mathcal{P}(X = 2) + \mathcal{P}(X = 3) + \mathcal{P}(X = 4)$$

$$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5}.$$

2. We are given that $X \sim \text{dunif}(a, b)$ so, we know that

$$E(X) = \frac{a+b}{2} \text{ and } SD(X) = \sigma = \sqrt{\frac{(b-a+1)^2 - 1}{12}}$$

According to the question, we have

$$\frac{a+b}{2} = 11 \text{ and } SD(X) = \sigma = \sqrt{\frac{(b-a+1)^2 - 1}{12}} = \sqrt{2}$$

$\Rightarrow a + b = 22$ ... (9.103)

and $(b-a+1)^2 - 1 = 24 \Rightarrow (b-a+1)^2 = 25 \Rightarrow b - a + 1 = \pm 5$

$$\Rightarrow \begin{cases} b - a = 4 \\ b - a = -6 \end{cases}$$

But we know that $b \geq a$ always. So, rejecting $b - a = -6$, we have

$\Rightarrow b - a = 4$ ... (9.104)

Adding (9.103) and (9.104), we get

$2b = 26 \Rightarrow b = 13.$

Putting b = 13 in (9.103), we get

$a + 13 = 22 \Rightarrow a = 22 - 13 \Rightarrow a = 9.$

Hence, the parameters of the discrete uniform distribution are a = 9 and b = 13.

3. We are given that $X \sim \text{Bern}(p)$ so, we know that

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

According to the question, we have

$$SD(X) = \sigma = \sqrt{p(1-p)} = \sqrt{0.25}$$

$$\Rightarrow p(1-p) = 0.25 \Rightarrow p - p^2 = \frac{1}{4} \Rightarrow 4p - 4p^2 = 1 \Rightarrow 4p^2 - 4p + 1 = 0$$

$$\Rightarrow (2p-1)^2 = 0 \Rightarrow 2p - 1 = 0 \Rightarrow p = \frac{1}{2}.$$

Hence, the parameter of the Bernoulli distribution is 1/2.

Reference

- Peter H. Westfall (2014) Kurtosis as Peakedness, 1905–2014. R.I.P., The American Statistician, 68:3, 191-195,

  DOI: 10.1080/00031305.2014.917055

UNIT 10

BINOMIAL AND MULTINOMIAL DISTRIBUTIONS

Structure

10.1 INTRODUCTION

In Unit 9, you have studied what is Bernoulli trial. In the same unit, you have studied Bernoulli distribution which was based on a single Bernoulli trial. If instead of a single trial, we have 'n' a finite number of Bernoulli trials and we are interested in the probability distribution of the number of successes in n Bernoulli trials then the name of the probability distribution which is developed for such situations is known as a binomial distribution. In Sec. 10.2, we will discuss its PMF and CDF while in Sec. 10.3, we will discuss MGF and some other summary measures of this probability distribution.

In Secs. 10.4 and 10.5, we will do similar studies about multinomial distribution. Some applications of some measures of these distributions are discussed in Sec. 10.6.

What we have discussed in this unit is summarised in Sec. 10.7. Self-Assessment Questions (SAQs) have been given in some sections which are generally based on the content discussed in that section. But to give you a good practice of what we have discussed in this unit some more questions based on the entire unit are given in Sec. 10.8 under the heading Terminal Questions. Due to the reason mentioned in Sec. 1.1 of Unit 1 of this course, solutions of all the SAQs and Terminal Questions are given in Sec. 10.9.

In the next unit, you will do a similar study about two more discrete probability distributions known as Poisson and hypergeometric distributions.

Expected Learning Outcomes

After completing this unit, you should be able to:

- ❖ identify situations where you can apply binomial and multinomial distributions;
- ❖ define PMF, CDF, MGF and some summary measures of binomial and multinomial distributions; and
- ❖ apply binomial and multinomial distributions to solve problems based on these two probability distributions.

10.2 STORY, DEFINITION, PMF AND CDF OF BINOMIAL DISTRIBUTION

Like Sec. 9.2 of the previous unit in this section, we will discuss one more special discrete distribution known as binomial distribution. In this section, we will also discuss the PMF and CDF of the binomial distribution. Recall that in Bernoulli distribution we have only a single Bernoulli trial but in binomial distribution, we have ‘n’, a finite number of Bernoulli trials. Before moving further first of all you should know the answers to the following two important questions.

- Are the ‘n’ trials independent or dependent? … (10.1)
- Does the probability of success in each trial remain the same or may vary? … (10.2)

Answers to these questions will decide which distribution is suitable. If we want to apply binomial distribution then answers to these two questions should be as follows.

- All the ‘n’ trials should be independent. … (10.3)
- The probability of success in each trial should remain the same. (10.4)

Now, we can discuss the story of the binomial distribution.

Story of Binomial Distribution: If we are performing a random experiment n-time and the outcome/realisation of each trial has only two categories traditionally known as success or failure such that:

- all these n Bernoulli trials are independent of each other and
- the probability of success (p) in each trial should be the same,

then the random variable (X) which counts the number of successes in these n Bernoulli trials follows a binomial distribution with parameters n and p and is denoted by X ~ B(n, p). … (10.5)

Now recall (9.15) in all discrete probability distributions $\mathcal{F}$ of the probability triplet $(\Omega, \mathcal{F}, \mathcal{P})$ is always the power set of Ω. So, out of three things $\Omega, \mathcal{F}, \mathcal{P}$ one thing $\mathcal{F}$ has been fixed for all the discrete probability distributions. The remaining two things Ω and $\mathcal{P}$ will vary from distribution to distribution. So, the moral of the story is as soon as we specify Ω and $\mathcal{P}$ then probability distribution is automatically specified. So, keeping this in view let us define binomial distribution as follows.

Definition and PMF of Binomial Distribution: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \{\omega_1\omega_2\omega_3 \ldots \omega_n : \text{ each } \omega_i = \text{success or failure}\}$ contains all possible 2^n sequences of success and failures of length n. If we define the random variable X on the sample space Ω by

$$X(\omega_1\omega_2\omega_3 \ldots \omega_n) = \text{Number of i such that } \omega_i = \text{success}$$

then the random variable X may take values 0, 1, 2, 3, 4, …, n. Here X = 0 means all the n outcomes are failures and X = n means all the outcomes are successes, etc. We say that the random variable X follows binomial distribution if the probability measure $\mathcal{P}$ is defined by

$$\mathcal{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ if } x = 0, 1, 2, 3, 4, \ldots, n \qquad \ldots (10.6)$$

$$\text{where } 0 \le p \le 1 \text{ and } \binom{n}{x} = \frac{\lfloor n}{\lfloor x \lfloor n-x} = \text{binomial coefficient} \qquad \ldots (10.7)$$

So, PMF of binomial distribution is given by

$$p_X(x) = \mathcal{P}(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x = 0, 1, 2, 3, 4, \ldots, n \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (10.8)$$

If random variable X follows binomial distribution with n independent Bernoulli trials and the probability of success p in each trial is constant, then n and p are known as parameters of the binomial distribution and is denoted by writing X ~ Bin(n, p). … (10.9)

Like the Bernoulli distribution case, we read X ~ Bin(n, p) as X is distributed as a binomial distribution with parameters n and p. Or we read it as X follows a binomial distribution with parameters n and p. … (10.10)

Since the statistical software used for hands-on practice for the lab courses of this programme is R programming language, so you should also know its notation in R. In R notation that is used for binomial distribution is binom(size, prob) in the stats package, where size represents the value of n and prob represents the value of p. In fact, like any probability distribution there are four function for binomial distribution namely dbinom(x, size, prob, …), pbinom(q, size, prob, …), qbinom(p, size, prob, …) and rbinom(n, size, prob, …). We have already explained meaning of these functions in Unit 9. ... (10.11)

Let us check the **validity of the PMF of the binomial distribution**.

We have to check two conditions non-negativity and normality refer to (4.62).

(1) **Non-negativity**: Since $\binom{n}{x} > 0, \; p \ge 0,$ so

$$\binom{n}{x} p^x (1-p)^{n-x} \ge 0, \; \forall \; x = 0, 1, 2, 3, 4, \ldots, n \qquad \ldots (10.12)$$

(2) **Normality**: Binomial theorem which you have studied in school mathematics states that:

$$(a+b)^n = \binom{n}{0}a^n b^0 + \binom{n}{1}a^{n-1}b^1 + \binom{n}{2}a^{n-2}b^2 + \ldots + \binom{n}{n}a^0 b^n$$

$$= \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k \qquad \ldots\ (10.13)$$

Replacing k by x, a by 1 – p, b by p in (10.13), we get

$$(1-p+p)^n = \sum_{x=0}^{n}\binom{n}{x}p^x(1-p)^{n-x} \Rightarrow (1)^n = \sum_{x=0}^{n}\binom{n}{x}p^x(1-p)^{n-x}$$

$$\Rightarrow \sum_{x=0}^{n}\binom{n}{x}p^x(1-p)^{n-x} = 1$$

This proves that sum of all probabilities of binomial distribution is 1.

Hence, we can say that PMF of the random variable which counts the number of successes in binomial distribution is a valid PMF.

One question that will be arising in your mind is how the expression $\binom{n}{x}p^x(1-p)^{n-x}$ comes in the definition of probability measure given by (10.6).

This is really a very good question such types of questions are necessary for learning the subject the way a master's degree learner should learn. Let us explain it as follows.

Let S and F denote the success and the failure respectively. Let X count the number of successes in n trials. Let $\underbrace{SSFFFSFSFFSSSF\ldots FS}_{n\text{-times}}$ be a particular sequence of n trials with x successes and n – x failures. Let p and q = 1 – p denote the probability of success and failure respectively. Therefore,

$$\mathcal{P}(S) = p, \text{ and } \mathcal{P}(F) = q = 1-p \qquad \ldots\ (10.14)$$

Now, probability of this particular sequence of success and failure is

$$\mathcal{P}\left(\underbrace{SSFFFSFSFF\ldots FS}_{n\text{-times}}\right) = \mathcal{P}(S)\mathcal{P}(S)\mathcal{P}(F)\mathcal{P}(F)\mathcal{P}(F)$$

$$\mathcal{P}(S)\mathcal{P}(F)\mathcal{P}(S)\mathcal{P}(F)\mathcal{P}(F)\ldots\mathcal{P}(F)\mathcal{P}(S) \left[\begin{array}{l}\text{If E and F are independent}\\ \text{events, then}\\ \mathcal{P}(E\cap F) = \mathcal{P}(E)\mathcal{P}(F)\end{array}\right]$$

$$= ppqqqpqpqq\ldots qp \qquad [\text{Using (10.14)}]$$

$$= \underbrace{ppp\ldots p}_{x\text{-times}}\underbrace{qqq\ldots q}_{(n-x)\text{-times}}$$

$$= p^x q^{n-x} \qquad \ldots\ (10.15)$$

Probability given by (10.15) is the probability of one particular sequence having x successes and n – x failures. But from the concept of combination studied in school mathematics, you know that out of n positions x positions can be filled up with successes and remaining n – x positions with failures in $\binom{n}{x}$ number of ways. Hence,

$\mathcal{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, if $x = 0, 1, 2, 3, 4, \ldots, n$. This completes the explanation of the answer of your question. … (10.16)

Now, we define the CDF of binomial distribution.

CDF of Binomial Distribution: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \{\omega_1\omega_2\omega_3 \ldots \omega_n : \text{ each } \omega_i = \text{success or failure}\}$ contains all possible 2^n sequences of success and failures of length n. If we define the random variable X on the sample space Ω by

$$X(\omega_1\omega_2\omega_3 \ldots \omega_n) = \text{Number of i such that } \omega_i = \text{success}$$

then the random variable X may take values 0, 1, 2, 3, 4, …, n. We say that the random variable X follows binomial distribution if probability measure $\mathcal{P}$ is defined by

$$\mathcal{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ if } x = 0, 1, 2, 3, 4, \ldots, n$$

where $0 \le p \le 1$

So, PMF of binomial distribution is given by

$$p_X(x) = \mathcal{P}(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x = 0, 1, 2, 3, 4, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}(X \le x) = \begin{cases} 0, & \text{if } x < 0 \\ \sum_{k=1}^{[x]} \binom{n}{k} p^k (1-p)^{n-k}, & \text{if } 0 \le x < n \\ 1, & \text{if } x \ge n \end{cases} \quad \ldots (10.17)$$

Let us do one example.

Example 1: Plot PMF of binomial random variable where n = 8 and p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9.

Solution: Let us first obtain $\mathcal{P}(X = x)$ for x = 0, 1, 2, 3, 4, 5, 6, 7 and 8 when n = 8 and p = 0.1 using PMF of binomial Bin(8, 0.1).

We know that $p_X(x) = \mathcal{P}(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x = 0, 1, 2, 3, 4, \ldots, n \\ 0, & \text{otherwise} \end{cases}$

In our case n = 8, p = 0.1, so we have

$$\left.\begin{array}{l}
\mathcal{P}(X=0)=\binom{8}{0}(0.1)^{0}(1-0.1)^{8-0}=0.43046721\\
\text{Similarly, } \mathcal{P}(X=1)=\binom{8}{1}(0.1)^{1}(1-0.1)^{8-1}=0.38263752\\
\mathcal{P}(X=2)=\binom{8}{2}(0.1)^{2}(1-0.1)^{8-2}=0.14880348\\
\mathcal{P}(X=3)=\binom{8}{3}(0.1)^{3}(1-0.1)^{8-3}=0.03306744\\
\mathcal{P}(X=4)=\binom{8}{4}(0.1)^{4}(1-0.1)^{8-4}=0.00459270\\
\mathcal{P}(X=5)=\binom{8}{5}(0.1)^{5}(1-0.1)^{8-5}=0.00040824\\
\mathcal{P}(X=6)=\binom{8}{6}(0.1)^{6}(1-0.1)^{8-6}=0.00002268\\
\mathcal{P}(X=7)=\binom{8}{7}(0.1)^{7}(1-0.1)^{8-7}=0.00000072\\
\mathcal{P}(X=8)=\binom{8}{8}(0.1)^{8}(1-0.1)^{8-8}=0.00000001
\end{array}\right\} \quad \ldots (10.18)$$

You can also obtain these probabilities with a single command in R as follows.

```
> dbinom(0:8,8,0.1)
[1] 0.43046721 0.38263752 0.14880348 0.03306744 0.00459270 0.00040824 0.00002268
[8] 0.00000072 0.00000001
```

Or we can specify names of the arguments. Remember when we do not specify names of the arguments then R matches them by their positions.

```
> dbinom(x=0:8, size = 8, prob = 0.1)
[1] 0.43046721 0.38263752 0.14880348 0.03306744 0.00459270 0.00040824 0.00002268
[8] 0.00000072 0.00000001
```

… (10.19)

Similarly, using R, we can obtain probabilities for other values of p as follows.

```
> dbinom(0:8,8,0.2)
[1] 0.16777216 0.33554432 0.29360128 0.14680064 0.04587520 0.00917504 0.00114688
[8] 0.00008192 0.00000256
> dbinom(0:8,8,0.3)
[1] 0.05764801 0.19765032 0.29647548 0.25412184 0.13613670 0.04667544 0.01000188
[8] 0.00122472 0.00006561
> dbinom(0:8,8,0.4)
[1] 0.01679616 0.08957952 0.20901888 0.27869184 0.23224320 0.12386304 0.04128768
[8] 0.00786432 0.00065536
> dbinom(0:8,8,0.5)
[1] 0.00390625 0.03125000 0.10937500 0.21875000 0.27343750 0.21875000 0.10937500
[8] 0.03125000 0.00390625
> dbinom(0:8,8,0.6)
[1] 0.00065536 0.00786432 0.04128768 0.12386304 0.23224320 0.27869184 0.20901888
[8] 0.08957952 0.01679616
> dbinom(0:8,8,0.7)
[1] 0.00006561 0.00122472 0.01000188 0.04667544 0.13613670 0.25412184 0.29647548
[8] 0.19765032 0.05764801
> dbinom(0:8,8,0.8)
[1] 0.00000256 0.00008192 0.00114688 0.00917504 0.04587520 0.14680064 0.29360128
[8] 0.33554432 0.16777216
> dbinom(0:8,8,0.9)
[1] 0.00000001 0.00000072 0.00002268 0.00040824 0.00459270 0.03306744 0.14880348
[8] 0.38263752 0.43046721
```

… (10.20)

Now, PMF of binomial random variable where n = 8 and p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9 are shown in Fig. 10.1 (a), (b), (c), (d), (e), (f), (g), (h) and (i) respectively given as follows.

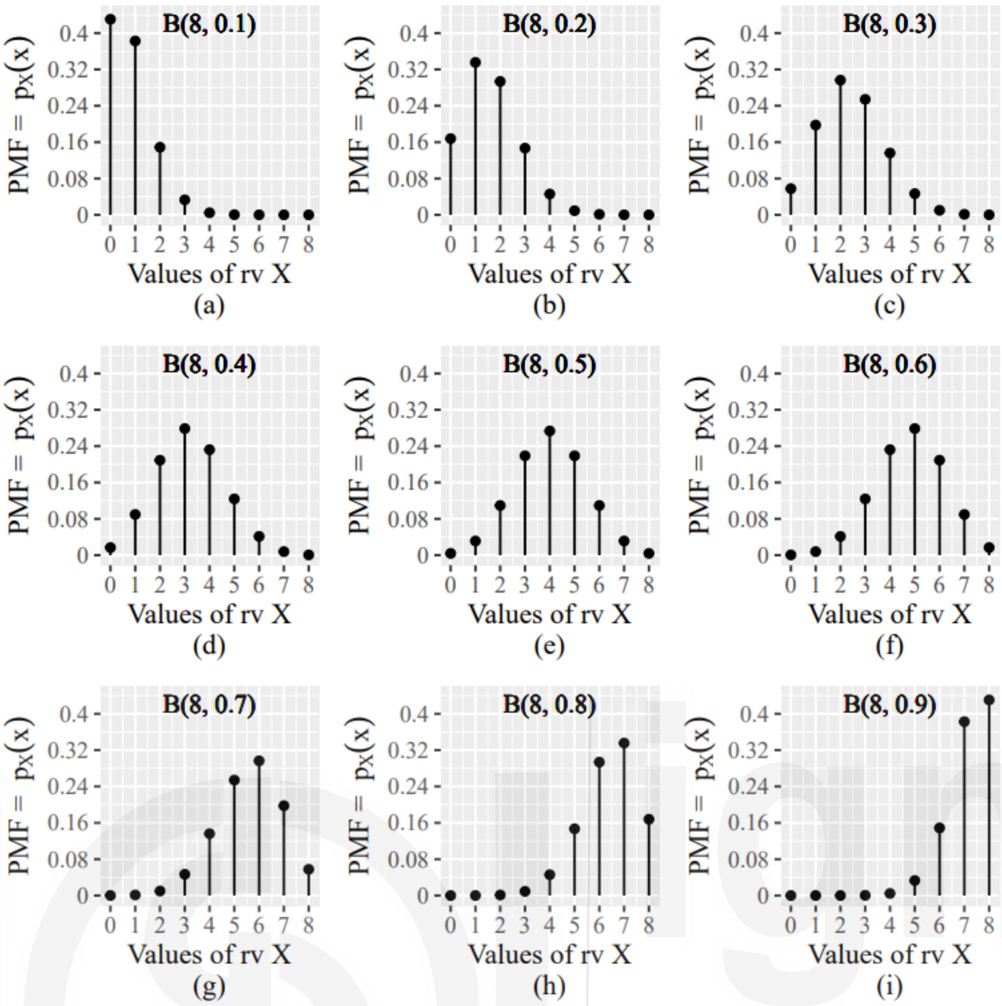


Fig. 10.1: Visualisation of PMF of binomial distributions where n = 8 and p is (a) 0.1 (b) 0.2 (c) 0.3 (d) 0.4 (e) 0.5 (f) 0.6 (g) 0.7 (h) 0.8 (i) 0.9

Example 2: A bag contains 4 red and 12 black balls. Three balls are drawn one by one with replacement. Plot PMF and CDF of number of red balls.

Solution: Let X denote the number of red balls drawn from the bag out of the three draws where balls are drawn one by one with replacement. So, X can take values 0, 1, 2 and 3. Since balls are drawn one by one with replacement so probability of getting a red ball in each draw is the same and is given by:

$$\mathcal{P}(\text{getting a red ball}) = \frac{4}{16} = \frac{1}{4} \left[\because \binom{4}{1} = 4 \text{ and } \binom{16}{1} = 16\right] \quad \ldots (10.21)$$

So, $X \sim B\left(3, \frac{1}{4}\right)$. To obtain PMF of X, first, we have to obtain probabilities X = 0, 1, 2 and 3 and are given by

$$\left.\begin{aligned}
\mathcal{P}(X=0) &= \binom{3}{0}\left(\frac{1}{4}\right)^0\left(1-\frac{1}{4}\right)^{3-0} = (1)(1)\left(\frac{3}{4}\right)^3 = \frac{27}{64} \\
\mathcal{P}(X=1) &= \binom{3}{1}\left(\frac{1}{4}\right)^1\left(1-\frac{1}{4}\right)^{3-1} = (3)\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^2 = \frac{27}{64} \\
\mathcal{P}(X=2) &= \binom{3}{2}\left(\frac{1}{4}\right)^2\left(1-\frac{1}{4}\right)^{3-2} = (3)\left(\frac{1}{16}\right)\left(\frac{3}{4}\right)^1 = \frac{9}{64} \\
\mathcal{P}(X=3) &= \binom{3}{3}\left(\frac{1}{4}\right)^3\left(1-\frac{1}{4}\right)^{3-3} = (1)\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^0 = \frac{1}{64}
\end{aligned}\right\} \quad \ldots (10.22)$$

Now, using (10.22) CDF of X is given by

$$F_X(x) = \mathcal{P}(X \le x) = \begin{cases} 0, & \text{if } x < 0 \\ \dfrac{27}{64}, & \text{if } 0 \le x < 1 \\ \dfrac{54}{64}, & \text{if } 1 \le x < 2 \\ \dfrac{63}{64}, & \text{if } 2 \le x < 3 \\ 1, & \text{if } x \ge 3 \end{cases} \qquad \ldots (10.23)$$

PMF and CDF are plotted in Fig. 10.2 (a) and (b) respectively as follows.

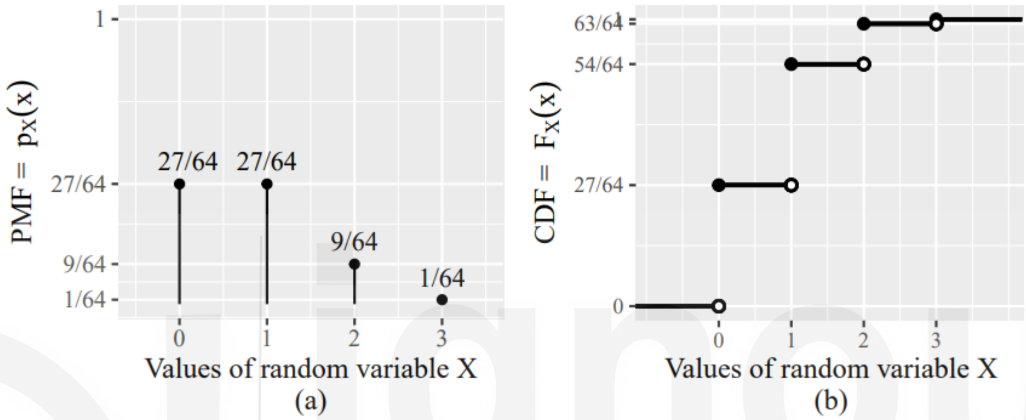


Fig. 10.2: Visualisation of (a) PMF (b) CDF of the B(3, 1/4) discussed in Example 2

Remark 2: In the special case p = 0, we say that X is a constant random variable having only one value 0 with probability 1. Similarly, in the other special case p = 1 we say that X is a constant random variable having only one value 1 with probability 1. … (10.24)

10.3 MGF AND OTHER SUMMARY MEASURES OF BINOMIAL DISTRIBUTION

In the previous section, you have studied PMF and CDF of binomial distribution. In this section we want to obtain MGF and some other summary measure of binomial distribution like mean, median, variance, etc. Let us first obtain MGF of binomial distribution. We will obtain MGF of binomial distribution using MGF of Bernoulli distribution. Note that if X ~ Bin(n, p) then X counts the number of successes in n Bernoulli trials. So, if I_i be the indicator random variable of success in i^{th} trial then $I_i \sim \text{Bern}(p)$ and therefore, using (9.54) or (9.55), we have

$$\text{MGF of Bern}(p) = M_X(t) = q + p\,e^t \qquad \ldots (10.25)$$

Using (9.61), we have

$$E(I_i) = p \text{ and } V(I_i) = p(1-p) \qquad \ldots (10.26)$$

Now, random variable X and the sum of the n indicator random variables $I_1 + I_2 + I_3 + \ldots + I_n$ both counts the number of successes in n trials, so we have

$$X = I_1 + I_2 + I_3 + \ldots + I_n \qquad \ldots (10.27)$$

$\therefore$ MGF of $X = M_X(t) = M_{I_1+I_2+I_3+\ldots+I_n}(t)$ $\quad$ [Using (10.27)]

$= M_{I_1}(t)\, M_{I_2}(t)\, M_{I_3}(t) \ldots M_{I_n}(t)$ $\quad$ [Using (7.94)]

$= \left(q+pe^t\right)\left(q+pe^t\right)\left(q+pe^t\right)\ldots\left(q+pe^t\right)$ $\quad$ [Using (9.55)]

$= \left(q+pe^t\right)^n$ $\quad$ ... (10.28)

Applying expectation on both sides of (10.27), we have

$E(X) = E\left(I_1 + I_2 + I_3 + \ldots + I_n\right)$

$= E\left(I_1\right) + E\left(I_2\right) + E\left(I_3\right) + \ldots + E\left(I_n\right)$ $\quad$ [Using addition theorem of expectation refer to (7.30)]

$= p + p + p + \ldots + p$ $\quad$ [Using (10.26)]

$= np$ $\quad$ ... (10.29)

Like expectation now after applying variance operator on both sides of (10.27), we have

$V(X) = V\left(I_1 + I_2 + I_3 + \ldots + I_n\right)$

$= V\left(I_1\right) + V\left(I_2\right) + V\left(I_3\right) + \ldots + V\left(I_n\right)$ $\quad$ [Using (7.74)]

$= p(1-p) + p(1-p) + p(1-p) + \ldots + p(1-p)$ $\quad$ [Using (10.26)]

$= np(1-p)$ $\quad$ ... (10.30)

We know that standard deviation of X is positive square root of variance of X. Hence, $SD(X) = \sqrt{\text{Variance of } X} = \sqrt{np(1-p)}$ $\quad$ … (10.31)

Let us obtain mode of the binomial distribution. We know that mode will be that value of X = x such that $p_X(x)$ is maximum. To get such a value we have to compare probabilities of values of the random variable X. We know that to compare two quantities either we take ratio or difference of the two quantities. Here ratio will be more suitable for our purpose. So, let do that.

$$\frac{p_X(x+1)}{p_X(x)} = \frac{\binom{n}{x+1} p^{x+1}(1-p)^{n-x-1}}{\binom{n}{x} p^x (1-p)^{n-x}} = \frac{\dfrac{\underline{|n}}{\underline{|x+1}\,\underline{|n-x-1}}\,p}{\dfrac{\underline{|n}}{\underline{|x}\,\underline{|n-x}}\,(1-p)} = \frac{(n-x)p}{(x+1)(1-p)} \quad \ldots (10.32)$$

Now, $\dfrac{p_X(x+1)}{p_X(x)} \geq 1 \Rightarrow \dfrac{(n-x)p}{(x+1)(1-p)} \geq 1 \Rightarrow np - xp \geq x - xp + 1 - p \Rightarrow np + p - 1 \geq x$

or $x \leq p(n+1) - 1$ $\quad$ … (10.33)

But x is an integer and $p(n+1)-1$ may not be an integer. However, to check whether $p(n+1)-1$ is an integer or not it is enough to check $p(n+1)$. So, two cases arise:

Case I: $p(n+1)$ is an integer. In this case, we have two modes $x = p(n+1)-1$ and $x+1 = p(n+1)$. For example, in Example 2 $p(n+1) = \frac{1}{4}(3+1) = \frac{4}{4} = 1$ which is an integer. So, we will have two modes $x = p(n+1) - 1 = \frac{1}{4}(3+1) - 1 = 0$ and $x + 1 = p(n+1) = \frac{1}{4}(3+1) = 1$. You can verify it from Fig. 10.2 where you see that $\mathcal{P}(X=0)=\mathcal{P}(X=1)=\frac{27}{64}$ which is maximum when we compare the probabilities of X = 0, 1, 2 and 3. Hence, binomial distribution discussed in Example 2 has two modes 0 and 1. (10.34)

Case II: $p(n+1)$ is not an integer. In this case, we have only one mode $x=[p(n+1)]$. For example, in Example 1 in all the nine binomial distributions value of $p(n+1)$ is not an integer. So, all the nine binomial distributions discussed in Example 1 have unique mode. You can verify it from Fig. 10.2 (a) to (i). For example, in the binomial distribution discussed in part (b) of Example 1 we see that $p(n+1)=(0.2)(8+1)=(0.2)(9)=1.8$ which is not an integer. So, we will have unique mode and this unique mode is $x=[p(n+1)]=[1.8]=1.$ So, mode of Bin(8, 0.2) is 1. Verify it by comparing height (probability) at the point X = 1 in Fig. 10.1 (b). If you want to recall greatest integer function, you can refer to (1.54) of the course MST-011. Similarly, in the binomial distribution Bin(8, 0.8) discussed in part (h) of Example 1, we see that $p(n+1)=(0.8)(8+1)=(0.8)(9)=7.2$ which is not an integer. So, we will have unique mode and this unique mode is $x=[p(n+1)]=[7.2]=7.$ So, mode of Bin(8, 0.8) is 7. Verify it by comparing height (probability) at the point x = 7 in Fig. 10.1 (h). … (10.35)

Similarly, we can obtain other measures of our interest. But keeping applied nature of the programme in view we are not focusing on proof of each measure. Some commonly used summary measures of binomial distribution are shown in Table 10.1 given as follows.

Table 10.1: Summary measures of binomial distribution

Name of measure	Formula	Name of measure	Formula
Mean	np	Standard deviation	$\sqrt{np(1-p)}$
Median	If np is an integer then median will be equal to mean np. Otherwise, median will not unique except some special cases.	MGF	$\left(q+pe^t\right)^n$
Mode	If $p(n+1)$ is an integer then there will be two modes $p(n+1)-1$ and $p(n+1)$ while if $p(n+1)$ is not an integer then there will be unique mode $[p(n+1)]$	Skewness	$\frac{1-2p}{\sqrt{npq}}$
Variance	$np(1-p)$	Kurtosis	$3+\frac{1-6pq}{np(1-p)}$

10.4 STORY, DEFINITION, PMF AND CDF OF MULTINOMIAL DISTRIBUTION

Recall that (refer 10.13 to 10.16) to prove normality of PMF of binomial distribution and to explain the involvement of the term $\binom{n}{x}p^x(1-p)^{n-x}$ in PMF of the binomial distribution, we took the help of binomial expansion which is known as binomial theorem. Similarly, to get similar understanding about multinomial distribution, we have to first understand multinomial expansion. We claim that multinomial expansion of $\left(a_1+a_2+a_3+\ldots+a_m\right)^n$ is given by

$$\left(a_1+a_2+a_3+\ldots+a_m\right)^n=\sum_{n_1,n_2\ldots n_m}\frac{\lfloor n}{\lfloor n_1\lfloor n_2\lfloor n_3\ldots\lfloor n_m}a_1^{n_1}a_2^{n_2}a_3^{n_3}\ldots a_m^{n_m} \qquad \ldots (10.36)$$

where $n_1+n_2+n_3+\ldots+n_m=n$ ... (10.37)

To understand (10.36) and (10.37) let us consider a particular expansion of binomial theorem as follows.

$$\left(a+b\right)^7=\binom{7}{0}a^7b^0+\binom{7}{1}a^6b+\binom{7}{2}a^5b^2+\binom{7}{3}a^4b^3+\binom{7}{4}a^3b^4+\binom{7}{5}a^2b^5$$
$$+\binom{7}{6}ab^6+\binom{7}{7}a^0b^7$$
$$=a^7b^0+7a^6b+21a^5b^2+35a^4b^3+35a^3b^4+21a^2b^5+7ab^6+a^0b^7 \quad (10.38)$$

Also, note that

$$\left(a+b\right)^7=\underbrace{(a+b)(a+b)(a+b)(a+b)(a+b)(a+b)(a+b)}_{7\text{ - times}} \qquad \ldots (10.39)$$

Now, let us consider any one term of RHS of (10.38) say third term $21a^5b^2$. In this term exponents of a and b are 5 and 2 respectively. Recall ordinary rule of multiplication which you have learnt in school mathematics. Keep that ordinary rule of multiplication in your mind. Each term of RHS of (10,38) is obtained by multiplying one factor (a or b) from each of the seven parentheses shown in RHS of (10.39) using ordinary rule of multiplication. To obtain a^5b^2 forget about its coefficient 21 for some time, think how many times you have to pick factor a out of the seven parentheses in RHS of (10.39) and how many times factor b. Obviously, you have to pick 5 times factor a and 2 times factor b. But out of 7 parentheses you can select 5 of them in $\binom{7}{5}$ ways. After doing so automatically, we will select b from the remaining two parentheses. So, finally, you will get

$$\underbrace{aaaaabb+aaaaabb+aaaaabb+\ldots+aaaaabb}_{\binom{7}{5}\text{- times}=21\text{ times}}$$
$$=\underbrace{a^5b^2+a^5b^2+a^5b^2+\ldots+a^5b^2}_{21\text{ times}}=21a^5b^2 \qquad \ldots (10.40)$$

Now, we apply this argument to understand the multinomial expansion given by (10.36). To obtain the general term $\frac{\lfloor n}{\lfloor n_1\lfloor n_2\lfloor n_3\ldots\lfloor n_m}a_1^{n_1}a_2^{n_2}a_3^{n_3}\ldots a_m^{n_m}$ of this expansion, we have to pick a_1 from n_1 parentheses out of n parentheses each containing m factors $a_1, a_2, a_3, \ldots, a_m$. Similarly, we will pick a_2 from the remaining $n-n_1$ parentheses; a_3 from the remaining $n-n_1-n_2$ parentheses and so on finally, a_m will be picked from the remaining $n-n_1-n_2-n_3-\ldots-n_{m-1}$ parentheses. This can be done in

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\ldots\binom{n-n_1-n_2-n_3-\ldots-n_{m-1}}{n_m} \text{ ways}$$

$$=\frac{\lfloor n}{\lfloor n_1\lfloor n-n_1}\times\frac{\lfloor n-n_1}{\lfloor n_2\lfloor n-n_1-n_2}\times\frac{\lfloor n-n_1-n_2}{\lfloor n_3\lfloor n-n_1-n_2-n_3}\times\ldots\times\frac{\lfloor n-n_1-n_2-n_3-\ldots-n_{m-1}}{\lfloor n_m\lfloor n-n_1-n_2-\ldots-n_{m-1}-n_m}$$

$$=\frac{\lfloor n}{\lfloor n_1\lfloor n_2\lfloor n_3\lfloor n_4\ldots\lfloor n_m} \qquad [\text{All other factors cancel out in pairs}]$$

Hence, $\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\ldots\binom{n-n_1-n_2-n_3-\ldots-n_{m-1}}{n_m}=\frac{\lfloor n}{\lfloor n_1\lfloor n_2\lfloor n_3\ldots\lfloor n_m}$

This completes the explanation of general term of the multinomial expansion.

Last point regarding multinomial expansion which will be required to prove normality of multinomial distribution is if $a_1+a_2+a_3+\ldots+a_m=1$ then from (10.36), we have

$$(1)^n=\sum_{n_1,\,n_2\ldots n_m}\frac{\lfloor n}{\lfloor n_1\lfloor n_2\lfloor n_3\ldots\lfloor n_m}a_1^{n_1}\,a_2^{n_2}\,a_3^{n_3}\ldots a_m^{n_m} \qquad \ldots (10.41)$$

$$\Rightarrow\sum_{n_1,\,n_2\ldots n_m}\frac{\lfloor n}{\lfloor n_1\lfloor n_2\lfloor n_3\ldots\lfloor n_m}a_1^{n_1}\,a_2^{n_2}\,a_3^{n_3}\ldots a_m^{n_m}=1 \qquad \ldots (10.42)$$

Now, we can discuss multinomial distribution which is done as follows.

Recall that in Bernoulli distribution, we have **one trial** and **two categories of outcomes** of the trial traditionally known as success and failure. In binomial distribution, we have **n trials** and again **two categories of outcomes** of the trial like Bernoulli distribution traditionally known as success and failure. Now, another possibility is if we have a random experiment like binomial having n trials but unlike binomial each trial has k (> 2) categories of outcomes where probability of success of each of the k categories of outcomes in each of the n trials remains unchanged. That is, if we denote k categories of outcomes by k events E_i, $i=1, 2, 3, \ldots, k,$ where

$E_i=$ event that outcome of the trial falls in i^{th} category, $i=1, 2, 3, \ldots, k.$

Let $\mathcal{P}(E_i)=p_i,\ i=1, 2, 3, \ldots, k.$ Suppose out of n trials, events $E_1, E_2, E_3, \ldots, E_k$ occur $x_1, x_2, x_3, \ldots, x_k$ times respectively, where $x_1+x_2+x_3+\ldots+x_k=n.$

Now, we can write the story of multinomial distribution as follows.

Story of Multinomial Distribution: If we perform a random experiment and the realisation of each trial have possibility of k (> 2) outcomes/results/categories specified by k mutually exclusive and exhaustive events E_i, $i=1, 2, 3, \ldots, k,$ with $\mathcal{P}(E_i)=p_i,\ i=1, 2, 3, \ldots, k.$ Suppose out of n trials, events $E_1, E_2, E_3, \ldots, E_k$ occur $x_1, x_2, x_3, \ldots, x_k$ times respectively, where $x_1+x_2+x_3+\ldots+x_k=n.$ Then multinomial distribution requires following things to be a suitable candidate to model the situation.

(a) Like binomial distribution n should be a fixed finite number.

(b) Number of outcomes or categories or possibilities for each trial should be fixed and a finite number greater than 2. That is k > 2 should be fixed and finite.

(c) Trials should be independent.

(d) Each $\mathcal{P}\left(E_i\right) = p_i,\ i = 1, 2, 3, \ldots, k$ should remain the same in all trials. That is $\mathcal{P}\left(E_i\right) = p_i,\ i = 1, 2, 3, \ldots, k$ cannot change trial to trial and k > 2.

… (10.43)

Now, we define multinomial distribution as follows.

Definition and PMF of Bernoulli Distribution: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \{\omega_1\omega_2\omega_3 \ldots \omega_n :\ \text{each } \omega_i = C_1 \text{ or } C_2 \text{ or } C_3 \text{ or } \ldots \text{or } C_k\}$ contains all possible k^n sequences of $C_1, C_2, C_3, \ldots, C_k$ of length n. If we define k random variables $X_i, 1 \le i \le k$ on the sample space Ω by

$$X_i\left(\omega_1\omega_2\omega_3 \ldots \omega_n\right) = \text{Number of } j \text{ such that } \omega_j = C_i$$

So, each random variable $X_i, 1 \le i \le k$ may take values 0, 1, 2, 3, 4, …, n and counts the number of times outcomes favours category C_i or event E_i. We say that the random variables $X_i, 1 \le i \le k$ follows multinomial distribution if the probability measure $\mathcal{P}$ is defined by

$$\mathcal{P}\left(X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_k = x_k\right) = \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \lfloor x_3 \cdots \lfloor x_k} p_1^{x_1} p_2^{x_2} p_3^{x_3} \ldots p_k^{x_k},$$

where $x_i = 0, 1, 2, 3, 4, \ldots, n;\ 1 \le i \le k$ and $x_1 + x_2 + x_3 + \ldots + x_k = n$

… (10.44)

So, PMF, $p_X(x)$ of multinomial distribution is given by

$$\mathcal{P}\left(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k\right) = \begin{cases} \dfrac{\lfloor n}{\lfloor x_1 \lfloor x_2 \cdots \lfloor x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}, & \text{if } x_i = 0, 1, 2, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

… (10.45)

If random variables $X_i, 1 \le i \le k$ follow multinomial distribution with n independent trials each having outcome among the k categories and the probability of each category is $p_i, 1 \le i \le k$ in each trial and is constant, then n and $p_i, 1 \le i \le k$ are known as parameters of the multinomial distribution and is denoted by writing $(X_1, X_2, X_3, \ldots, X_k) \sim \text{multinom}(n; p_1, p_2, p_3, \ldots, p_k)$.

… (10.46)

We read it as $(X_1, X_2, X_3, \ldots, X_k)$ follows multinomial distribution with parameters $(n; p_1, p_2, p_3, \ldots, p_k)$. … (10.47)

Since the statistical software used for hands-on practice for the lab courses of this programme is R programming language, so you should also know its notation in R. In R notation that is used for multinomial distribution is multinom(size, prob) in the stats package, where size represents the value of n and prob represents the vector of probabilities $(p_1, p_2, p_3, \ldots, p_k)$. Here instead of four, we have two functions for multinomial distribution namely dmultinom(x, size, prob, …) and rmultinom(n, size, prob, …). We have already explained meaning of these functions in Unit 9. ... (10.48)

Let us check the **validity of the PMF of the multinomial distribution**.

We have to check two conditions non-negativity and normality refer to (4.62).

(1) **Non-negativity**: Since $\frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \cdots \lfloor x_k} > 0$ and $p_i^{x_i} \geq 0$, for each i, $1 \leq i \leq k$ so

$$\frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \cdots \lfloor x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \geq 0, \ \forall \ x_i = 0, 1, 2, 3, 4, \ldots, n, \ 1 \leq i \leq n \qquad \ldots (10.49)$$

(2) **Normality**: Using (10.42), we have

$$\sum_{x_1, x_2 \ldots x_k} \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \lfloor x_3 \cdots \lfloor x_m} x_1^{x_1} x_2^{x_2} x_3^{x_3} \ldots x_k^{x_k} = 1 \qquad \ldots (10.50)$$

This proves that sum of all probabilities of multinomial distribution is 1.

Hence, PMF defined by (10.45) of multinomial distribution is a valid PMF.

Now, we define CDF of multinomial distribution.

CDF of Multinomial Distribution: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \{\omega_1\omega_2\omega_3 \ldots \omega_n : \text{ each } \omega_i = C_1 \text{ or } C_2 \text{ or } C_3 \text{ or } \ldots \text{or } C_k\}$ contains all possible k^n sequences of $C_1, C_2, C_3, \ldots, C_k$ of length n. If we define k random variables $X_i, 1 \leq i \leq k$ on the sample space Ω by

$$X_i(\omega_1\omega_2\omega_3 \ldots \omega_n) = \text{Number of j such that } \omega_j = C_i$$

So, each random variable $X_i, 1 \leq i \leq k$ may take values 0, 1, 2, 3, 4, …, n and counts the number of times outcomes favours category C_i or event E_i. We say that the random variables $X_i, 1 \leq i \leq k$ follows multinomial distribution if PMF is given by

$$\mathcal{P}(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \begin{cases} \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \cdots \lfloor x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}, & \text{if } x_i = 0, 1, 2, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

where $x_i = 0, 1, 2, 3, 4, \ldots, n; 1 \leq i \leq k$ and $x_1 + x_2 + x_3 + \ldots + x_k = n$

Let $x_1, x_2, x_3, \ldots, x_k$ be any fixed k real numbers then CDF of the multinomial distribution random variable $(X_1, X_2, X_3, \ldots, X_k)$ is given by

$$\mathcal{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_k \leq x_k) = \sum_{x_1=0}^{[x_1]} \sum_{x_2=0}^{[x_2]} \cdots \sum_{x_k=0}^{[x_k]} \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \cdots \lfloor x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \qquad (10.51)$$

10.5 MGF AND OTHER SUMMARY MEASURES OF MULTINOMIAL DISTRIBUTION

In the previous section, you have studied PMF and CDF of multinomial distribution. In this section, we want to obtain MGF and some other summary measure of multinomial distribution like mean and variance. Let us first obtain MGF of multinomial distribution.

Calculation of MGF

Let $X = (X_1, X_2, X_3, \ldots, X_k)$ and $t = (t_1, t_2, t_3, \ldots, t_k)$, $t_i \in \mathbb{R}, 1 \le i \le k$ then using notations mentioned in (6.33) to (6.35), we have

$$t'X = \begin{bmatrix} t_1 & t_2 & \ldots & t_k \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{bmatrix} = t_1 X_1 + t_2 X_2 + \ldots + t_k X_k \qquad \ldots (10.52)$$

$$M_{(X_1, X_2, \ldots, X_k)}(t_1, t_2, \ldots, t_k) = M_X(t) = E\left(e^{t'X}\right) = E\left(e^{t_1 X_1 + t_2 X_2 + \ldots + t_k X_k}\right) \quad [\text{Using } (10.52)]$$

$$= \sum_{\substack{x_1, x_2, \ldots, x_k \\ x_1 + x_2 + \ldots + x_k = n}} e^{t_1 x_1 + t_2 x_2 + \ldots + t_k x_k} \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \ldots \lfloor x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

$$= \sum_{\substack{x_1, x_2, \ldots, x_k \\ x_1 + x_2 + \ldots + x_k = n}} e^{t_1 x_1} e^{t_2 x_2} e^{t_3 x_3} \ldots e^{t_k x_k} \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \ldots \lfloor x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

$$= \sum_{\substack{x_1, x_2, \ldots, x_k \\ x_1 + x_2 + \ldots + x_k = n}} \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \ldots \lfloor x_k} \left(e^{t_1} p_1\right)^{x_1} \left(e^{t_2} p_2\right)^{x_2} \ldots \left(e^{t_k} p_k\right)^{x_k}$$

$$= \left(e^{t_1} p_1 + e^{t_2} p_2 + \ldots + e^{t_k} p_k\right)^n \qquad [\text{Using } (10.36) \text{ and } (10.37)] \quad \ldots (10.53)$$

Now, MGF of X_i, $1 \le i \le k$ can be obtained by putting $t_j = 0, \ \forall \ j \ne i$

$$M_{X_i}(t_i) = \left(e^{t_1} p_1 + e^{t_2} p_2 + \ldots + e^{t_{i-1}} p_{i-1} + e^{t_i} p_i + e^{t_{i+1}} p_{i+1} + \ldots + e^{t_k} p_k\right)^n \Big|_{t_j = 0,\, j \ne i}$$

$$= \left(e^0 p_1 + e^0 p_2 + \ldots + e^0 p_{i-1} + e^{t_i} p_i + e^0 p_{i+1} + \ldots + e^0 p_k\right)^n$$

$$= \left(p_1 + p_2 + \ldots + p_{i-1} + e^{t_i} p_i + p_{i+1} + \ldots + p_k\right)^n$$

$$= \left(1 - p_i + e^{t_i} p_i\right)^n \qquad \begin{bmatrix} \because p_1 + p_2 + \ldots + p_{i-1} + p_i + p_{i+1} + \ldots + p_k = 1 \\ \Rightarrow p_1 + p_2 + \ldots + p_{i-1} + p_{i+1} + \ldots + p_k = 1 - p_i \end{bmatrix}$$

$$M_{X_i}(t_i) = \left(1 - p_i + e^{t_i} p_i\right)^n, \ i = 1, 2, 3, \ldots, k \qquad \ldots (10.54)$$

But it is MGF of binomial distribution with parameters n and p_i. Therefore by uniqueness theorem of MGF refer (7.95), we have

$$X_i \sim Bin(n, p_i), \ i = 1, 2, 3, \ldots, k \qquad \ldots (10.55)$$

Using (10.29) and (10.30), we have

$$E(X_i) = np_i, \quad V(X_i) = np_i(1 - p_i), \ i = 1, 2, 3, \ldots, k \qquad \ldots (10.56)$$

We know that standard deviation of a random variable is positive square root of variance of the random variable. Hence, standard deviation of X_i

$$= SD(X_i) = \sqrt{\text{Variance of } X_i} = \sqrt{np_i(1 - p_i)} \qquad \ldots (10.57)$$

Let us also obtain expected value of X_i, $i = 1, 2, 3, \ldots, k$ using extension of the techniques used in (7.57) to (7.60).

$$E(X_i) = \frac{\partial}{\partial x_i}\left(M_{(X_1, X_2, \ldots, X_k)}(t_1, t_2, \ldots, t_k)\right)_{(t_1, t_2, \ldots, t_k)=(0, 0, \ldots, 0)}$$

$$= \left(\frac{\partial}{\partial x_i}\left(e^{t_1}p_1 + e^{t_2}p_2 + \ldots + e^{t_k}p_k\right)^n\right)_{(t_1, t_2, \ldots, t_k)=(0, 0, \ldots, 0)}$$

$$= n\left(\left(e^{t_1}p_1 + e^{t_2}p_2 + \ldots + e^{t_k}p_k\right)^{n-1} e^{t_i}p_i\right)_{(t_1, t_2, \ldots, t_k)=(0, 0, \ldots, 0)}$$

$$= np_i\left(p_1 + p_2 + \ldots + p_k\right)^{n-1} \quad \left[\because e^0 = 1\right]$$

$$= np_i \quad \left[\because p_1 + p_2 + \ldots + p_k = 1\right]$$

$$\Rightarrow E(X_i) = np_i \qquad i = 1, 2, 3, \ldots, k \qquad \ldots (10.58)$$

Note that it matches with (10.56).

Similarly, we can obtain expected value of product $X_i X_j \;\; \forall \; i, j = 1, 2, 3, \ldots, k$

$$E(X_i X_j) = \frac{\partial^2}{\partial x_i \partial x_j}\left(M_{(X_1, X_2, \ldots, X_k)}(t_1, t_2, \ldots, t_k)\right)_{(t_1, t_2, \ldots, t_k)=(0, 0, \ldots, 0)}$$

$$= \left(\frac{\partial^2}{\partial x_i \partial x_j}\left(e^{t_1}p_1 + e^{t_2}p_2 + \ldots + e^{t_k}p_k\right)^n\right)_{(t_1, t_2, \ldots, t_k)=(0, 0, \ldots, 0)}$$

$$= n\left(\frac{\partial}{\partial x_i}\left(e^{t_1}p_1 + e^{t_2}p_2 + \ldots + e^{t_k}p_k\right)^{n-1} e^{t_j}p_j\right)_{(t_1, t_2, \ldots, t_k)=(0, 0, \ldots, 0)}$$

$$= n(n-1)\left(\left(e^{t_1}p_1 + e^{t_2}p_2 + \ldots + e^{t_k}p_k\right)^{n-2} e^{t_i}p_i\, e^{t_j}p_j\right)_{(t_1, t_2, \ldots, t_k)=(0, 0, \ldots, 0)}$$

$$= n(n-1)p_i p_j\left(p_1 + p_2 + \ldots + p_k\right)^{n-2} \quad \left[\because e^0 = 1\right]$$

$$= n(n-1)p_i p_j \quad \left[\because p_1 + p_2 + \ldots + p_k = 1\right]$$

$$\Rightarrow E(X_i X_j) = n(n-1)p_i p_j, \;\; \forall \; i, j = 1, 2, 3, \ldots, k; \, i \neq j \qquad \ldots (10.59)$$

Now, using (7.67), we have

$$\mathrm{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) \qquad \ldots (10.60)$$

Using (10.58) and (10.59) in (10.60), we get

$$\mathrm{Cov}(X_i, X_j) = n(n-1)p_i p_j - (np_i)(np_j)$$

$$= n^2 p_i p_j - np_i p_j - n^2 p_i p_j$$

$$= -np_i p_j \qquad \ldots (10.61)$$

But $X = (X_1, X_2, X_3, \ldots, X_k)$ so, we have

$$E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_k) \end{bmatrix} = \begin{bmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{bmatrix} = n\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{bmatrix} \text{ and}$$

$$(X - E(X))(X - E(X))' = \begin{bmatrix} X_1 - E(X_1) \\ X_2 - E(X_2) \\ \vdots \\ X_k - E(X_k) \end{bmatrix}\left[X_1 - E(X_1) \quad X_2 - E(X_2) \quad \ldots \quad X_k - E(X_k)\right]$$

$$= \begin{bmatrix} \left(X_1 - E(X_1)\right)^2 & \left(X_1 - E(X_1)\right)\left(X_2 - E(X_2)\right) & \cdots & \left(X_1 - E(X_1)\right)\left(X_k - E(X_k)\right) \\ \left(X_2 - E(X_2)\right)\left(X_1 - E(X_1)\right) & \left(X_2 - E(X_2)\right)^2 & \cdots & \left(X_2 - E(X_2)\right)\left(X_k - E(X_k)\right) \\ \vdots & \vdots & \ddots & \vdots \\ \left(X_k - E(X_k)\right)\left(X_1 - E(X_1)\right) & \left(X_k - E(X_k)\right)\left(X_2 - E(X_2)\right) & \ldots & \left(X_k - E(X_k)\right)^2 \end{bmatrix}$$

… (10.62)

We know that refer to (7.62) and (7.66) of this course.

$$\text{Variance of the distribution} = \mu_2 = E\left(X - E\left(X\right)\right)^2 \text{ and} \qquad \ldots (10.63)$$

$$\text{Cov}(X, Y) = E\left(\left(X - E(X)\right)\left(Y - E(Y)\right)\right) \qquad \ldots (10.64)$$

Using (10.56), (10.61), (10.63) and (10.64) in (10.62) variance covariance matrix of the random variable $X = \left(X_1, X_2, X_3, \ldots, X_k\right)$ is given by

$$E\left(\left(X - E(X)\right)\left(X - E(X)\right)'\right) = \begin{bmatrix} np_1(1-p_1) & -np_1 p_2 & \cdots & -np_1 p_k \\ -np_2 p_1 & np_2(1-p_2) & \cdots & -np_2 p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_k p_1 & -np_k p_2 & \ldots & np_k(1-p_k) \end{bmatrix} \ldots (10.65)$$

10.6 APPLICATIONS AND ANALYSIS OF BINOMIAL AND MULTINOMIAL DISTRIBUTIONS

In this section, we will apply binomial and multinomial distributions to solve some problems where assumptions of these distributions make sense.

Example 3: Find mean, variance and standard deviation of the random variable X where X ~ Bin(100, 4/5).

Solution: We know that if X ~ Bin(n, p) then

Mean = E(X) = np, Variance of X = V(X) = $np(1-p)$ and SD(X) = $\sqrt{np(1-p)}$.

In the present case: n = 100, p = 4/5. So, we have

Mean = 100(4/5) = 80, Variance = 100(4/5)(1/5) = 16 and SD = $\sqrt{16} = 4$.

Example 4: If $X = \left(X_1, X_2, X_3\right) \sim \text{multinom}\left(20, 1/10, 3/10, 6/10\right)$. Find mean vector of X. Also, find variance covariance matric of the random variable X.

Solution: Comparing $\text{multinom}\left(20, 1/10, 3/10, 6/10\right)$ with $\text{multinom}\left(n, p_1, p_2, p_3\right)$, we get

$$n = 20,\ p_1 = 1/10,\ p_2 = 3/10,\ p_3 = 6/10.$$

We know that mean vector of the random variable $X = \left(X_1, X_2, X_3\right)$ is given by

$$\text{Mean vector of } X = E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ E(X_3) \end{bmatrix} = \begin{bmatrix} np_1 \\ np_2 \\ np_3 \end{bmatrix} = \begin{bmatrix} 20 \times \frac{1}{10} \\ 20 \times \frac{3}{10} \\ 20 \times \frac{6}{10} \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ 12 \end{bmatrix}.$$

We also know that variance covariance matrix of the random variable $X = (X_1, X_2, X_3)$ is given by

$$\begin{bmatrix} np_1(1-p_1) & -np_1p_2 & -np_1p_3 \\ -np_2p_1 & np_2(1-p_2) & -np_2p_3 \\ -np_3p_1 & -np_3p_2 & np_3(1-p_3) \end{bmatrix}$$

$$= \begin{bmatrix} 20\left(\frac{1}{10}\right)\left(1-\frac{1}{10}\right) & -20\left(\frac{1}{10}\right)\left(\frac{3}{10}\right) & -20\left(\frac{1}{10}\right)\left(\frac{6}{10}\right) \\ -20\left(\frac{3}{10}\right)\left(\frac{1}{10}\right) & 20\left(\frac{3}{10}\right)\left(1-\frac{3}{10}\right) & -20\left(\frac{3}{10}\right)\left(\frac{6}{10}\right) \\ -20\left(\frac{6}{10}\right)\left(\frac{1}{10}\right) & -20\left(\frac{6}{10}\right)\left(\frac{3}{10}\right) & 20\left(\frac{6}{10}\right)\left(1-\frac{6}{10}\right) \end{bmatrix}$$

$$= \begin{bmatrix} 1.8 & -0.6 & -1.2 \\ -0.6 & 4.2 & -3.6 \\ -1.2 & -3.6 & 4.8 \end{bmatrix}$$

Example 5: Three dice are thrown. Find the probability of getting two 1 and one 3 (a) using classical approach of probability theory discussed in Unit 1 of this course and (b) using multinomial distribution.

Solution: (a) Let us first obtain required probability using classical approach. We know that when three dice are thrown then total number of possible outcomes in this random experiment are $6 \times 6 \times 6 = 6^3 = 216.$ Let E be the event of getting two 1 and one 3 then $E = \{(1, 1, 3), (1, 3, 1), (3, 1, 1)\}.$

Using classical approach to probability theory required probability is given by

$$\mathcal{P}(E) = \frac{n(E)}{n(\Omega)} = \frac{3}{216} = \frac{1}{72}. \qquad \ldots (10.66)$$

(b) Now, let us obtain required probability using multinomial distribution. When a die is thrown then there are six possible outcomes 1 or 2 or 3 or 4 or 5 or 6. So, let $E_i,\ i = 1, 2, 3, 4, 5, 6$ be the events of getting 1, 2, 3, 4, 5, 6 respectively. Let $X_i,\ i = 1, 2, 3, 4, 5, 6$ be the random variables which denote the number of times events $E_i,\ i = 1, 2, 3, 4, 5, 6$ occur respectively. In usual notations, we are given n = 3 because three dice are thrown or we can say that a die is thrown thrice and $p_i = \mathcal{P}(E_i)$ so

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6};\ x_1 = 2,\ x_2 = 0,\ x_3 = 1,\ x_4 = 0,\ x_5 = 0,\ x_6 = 0.$$

where $x_i = 0, 1, 2, 3;\ 1 \le i \le 6$ and $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 3$

Now, using multinomial distribution required probability is given by

$$\mathcal{P}\left(X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_6 = x_6\right) = \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \ldots \lfloor x_6} p_1^{x_1} p_2^{x_2} \ldots p_6^{x_6}$$

$$\therefore\ \mathcal{P}\left(X_1 = 2,\, X_2 = 0,\, X_3 = 1,\, X_4 = 0,\, X_1 = 0,\, X_2 = 0\right)$$

$$= \frac{\lfloor 3}{\lfloor 2 \lfloor 0 \lfloor 1 \lfloor 0 \lfloor 0 \lfloor 0}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^0 = \frac{6}{2}\left(\frac{1}{36}\right)\left(\frac{1}{6}\right) = \frac{1}{72} \quad \ldots (10.67)$$

From (10.66) and (10.67) we see that answers match as expected.

Now, you can try the following two Self-Assessment Questions.

SAQ 1

In a university 30% of the students stay in hostels of the university and remaining 70% commute from outside. If 15 students of this university are selected at random find the probability that exactly 4 of them stay in hostels.

SAQ 2

In a bag there are 5 red balls, 6 black balls, 3 blue balls and 6 yellow balls. Ten balls are drawn from this bag one by one with replacement. Find the probability that out of these 10 drawn balls 3 are red, 2 are black, 1 is blue and 4 are yellow.

10.7 SUMMARY

A brief summary of what we have covered in this unit is given as follows:

- **Definition of Binomial Distribution**: Let $(\Omega,\, \mathcal{F},\, \mathcal{P})$ be a probability space and $\Omega = \left\{\omega_1\omega_2\omega_3 \ldots \omega_n :\ \text{each } \omega_i = \text{success or failure}\right\}$ contains all possible 2^n sequences of success and failure of length n. If we define the random variable X on the sample space Ω by

  $$X\left(\omega_1\omega_2\omega_3 \ldots \omega_n\right) = \text{Number of i such that } \omega_i = \text{success}$$

  then the random variable X may take values 0, 1, 2, 3, 4, …, n. We say that the random variable X follows binomial distribution if the probability measure $\mathcal{P}$ is defined by

  $$\mathcal{P}\left(X = x\right) = \binom{n}{x} p^x (1-p)^{n-x},\ \text{if } x = 0, 1, 2, 3, 4, \ldots, n$$

  where $0 \le p \le 1$ and $\binom{n}{x} = \frac{\lfloor n}{\lfloor x \lfloor n-x} = \text{binomial coefficient}$

- PMF of binomial distribution is given by

  $$p_X(x) = \mathcal{P}\left(X = x\right) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x = 0, 1, 2, 3, 4, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

- **CDF of Binomial Distribution**: Let x be any fixed real number then CDF of the binomial random variable X is given by

$$F_X(x) = \mathcal{P}\left(X \le x\right) = \begin{cases} 0, & \text{if } x < 0 \\ \sum_{k=1}^{[x]} \binom{n}{k} p^k (1-p)^{n-k}, & \text{if } 0 \le x < n \\ 1, & \text{if } x \ge n \end{cases}$$

- **Summary measures of binomial distribution**

Name of measure	Formula	Name of measure	Formula
Mean	np	Standard deviation	$\sqrt{np(1-p)}$
Median	If np is an integer then median will be equal to mean np. Otherwise, median will not unique except some special cases.	MGF	$\left(q+pe^t\right)^n$
Mode	If $p(n+1)$ is an integer then there will be two modes $p(n+1)-1$ and $p(n+1)$ while if $p(n+1)$ is not an integer then there will be unique mode $\left[p(n+1)\right]$	Skewness	$\frac{1-2p}{\sqrt{npq}}$
Variance	$np(1-p)$	Kurtosis	$\frac{1-6pq}{np(1-p)}$

- **Definition of Multinomial Distribution**: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $\Omega = \left\{\omega_1\omega_2\omega_3 \ldots \omega_n : \text{ each } \omega_i = C_1 \text{ or } C_2 \text{ or } C_3 \text{ or } \ldots \text{or } C_k\right\}$ contains all possible k^n sequences of $C_1, C_2, C_3, \ldots, C_k$ of length n. If we define k random variables $X_i, 1 \le i \le k$ on the sample space Ω by

$$X_i\left(\omega_1\omega_2\omega_3 \ldots \omega_n\right) = \text{Number of } j \text{ such that } \omega_j = C_i$$

then each random variable $X_i, 1 \le i \le k$ may take values 0, 1, 2, 3, 4, …, n and counts the number of times outcomes favours category C_i or event E_i. We say that the random variables $X_i, 1 \le i \le k$ follows multinomial distribution if the probability measure $\mathcal{P}$ is defined by

$$\mathcal{P}\left(X_1 = x_1, X_2 = x_2, X_3 = x_3, \ldots, X_k = x_k\right) = \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \lfloor x_3 \ldots \lfloor x_k} p_1^{x_1} p_2^{x_2} p_3^{x_3} \ldots p_k^{x_k},$$

where $x_i = 0, 1, 2, 3, 4, \ldots, n; 1 \le i \le k$ and $x_1 + x_2 + x_3 + \ldots + x_k = n$

- PMF, $p_X(x)$ of multinomial distribution is given by

$$\mathcal{P}\left(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k\right) = \begin{cases} \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \cdots \lfloor x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}, & \text{if } x_i = 0, 1, 2, \ldots, n \\ 0, & \text{otherwise} \end{cases}$$

- **CDF of Multinomial Distribution**: Let $x_1, x_2, x_3, \ldots, x_k$ be any fixed k real numbers then CDF of the multinomial distribution random variable $\left(X_1, X_2, X_3, \ldots, X_k\right)$ is given by

$$\mathcal{P}\left(X_1 \le x_1, X_2 \le x_2, \ldots, X_k \le x_k\right) = \sum_{x_1=0}^{[x_1]} \sum_{x_2=0}^{[x_2]} \cdots \sum_{x_k=0}^{[x_k]} \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \cdots \lfloor x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

- MGF of Multinomial distribution is given by
  $M_{(X_1, X_2, \ldots, X_k)}(t_1, t_2, \ldots, t_k) = M_X(t) = \left(e^{t_1}p_1 + e^{t_2}p_2 + \ldots + e^{t_k}p_k\right)^n$

- Mean of the multinomial random variable $X = (X_1, X_2, X_3, \ldots, X_k)$ is given by $E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_k) \end{bmatrix} = \begin{bmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{bmatrix} = n\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{bmatrix}$

- Variance covariance matrix of multinomial random variable $X = (X_1, X_2, X_3, \ldots, X_k)$ is given by

$$\begin{bmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & -np_1p_k \\ -np_2p_1 & np_2(1-p_2) & \cdots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \ldots & np_k(1-p_k) \end{bmatrix}$$

10.8 TERMINAL QUESTIONS

1. If X ~ Bin(n, p) and Y ~ Bin(m, p) then find probability distribution of the random variable X + Y.
2. If in terminal question 1, we have m = 10, n = 8 and p = 0.5 then find $\mathcal{P}(X + Y = 4)$.
3. In a city probabilities of persons having blood types O, A, B and AB are 0.45, 0.4, 0.1 and 0.05 respectively. If 50 persons are selected from this city then find the probability that out of them 25, 20, 4 and 1 have blood types O, A, B and AB respectively.

10.9 SOLUTIONS/ANSWERS

Self-Assessment Questions (SAQs)

1. In this problem each student falls in one of the two categories stay in hostel and do not stay in hostel. Also, we have selected 15 students at random from this university. So, it is a situation of binomial distribution because if, we call stay in hostel as success and do not stay in hostel as failure then outcome of each trial is success or failure and, we have n = 15 trials. We are given that probability of success is p = 0.30 and so 1 – p = 0.70. Hence, required probability is given by

$$\mathcal{P}(X = 4) = \binom{15}{4}(0.3)^4(1-0.3)^{15-4}$$

$$= \frac{15.14.13.12}{4.3.2.1}(0.3)^4(0.7)^{11} = 0.2186231.$$

2. In this problem given bag contains balls of 4 colours. So, a drawn balls has 4 possibilities of colour. Also, balls are drawn one by one with replacement. It means trials are independent and probability of a getting a ball of a particular colour remains the same in each trial. Hence, it

satisfies all the requirements of multinomial distribution. If X_i, $i = 1, 2, 3, 4$ represents number of drawn balls of red, black, blue and yellow colours respectively, then in usual notations, we are given

$$p_1 = \frac{5}{20},\ p_2 = \frac{6}{20},\ p_3 = \frac{3}{20},\ p_4 = \frac{6}{20};\ x_1 = 3,\ x_2 = 2,\ x_3 = 1,\ x_4 = 4.$$

where $x_1 + x_2 + x_3 + x_4 = n = 10$

Now, using multinomial distribution required probability is given by

$$\mathcal{P}\left(X_1 = x_1,\ X_2 = x_2,\ X_3 = x_3,\ X_4 = x_4\right) = \frac{\lfloor n}{\lfloor x_1 \lfloor x_2 \lfloor x_3 \lfloor x_4} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}$$

$$\therefore \mathcal{P}\left(X_1 = 3,\ X_2 = 2,\ X_3 = 1,\ X_4 = 4\right) = \frac{\lfloor 10}{\lfloor 3 \lfloor 2 \lfloor 1 \lfloor 4}\left(\frac{5}{20}\right)^3\left(\frac{6}{20}\right)^2\left(\frac{3}{20}\right)^1\left(\frac{6}{20}\right)^4$$

$$= 0.02152828$$

Terminal Questions

1. Since X ~ Bin(n, p) it means the random variable X counts the number of successes out of n trials each having probability of success p. Similarly, Y ~ Bin(m, p) it means the random variable Y counts the number of successes out of m trials each having probability of success p. So, the random variable X + Y counts the number of successes out of m + n trials each having probability of success p. Hence, the random variable X + Y follows binomial distribution with parameters m + n and p. So, X + Y ~ Bin(m + n, p).

$$\mathcal{P}\left(X \ge 2\right) = \sum_{x=2}^{4} \mathcal{P}\left(X = x\right) = \mathcal{P}\left(X = 2\right) + \mathcal{P}\left(X = 3\right) + \mathcal{P}\left(X = 4\right)$$

$$= \frac{1}{5} + \frac{1}{5} + \frac{1}{5} = \frac{3}{5}.$$

2. We are given that m = 10, n = 8, p = 0.5, so, X + Y ~ Bin(18, 0.5). Hence, using PMF of binomial distribution, we have

$$\mathcal{P}\left(X + Y = 4\right) = \binom{18}{4}(0.5)^4 (1 - 0.5)^{18-4}$$

$$= \frac{18.17.16.15}{4.3.2.1}(0.5)^4 (0.5)^{14} = 0.01167297.$$

3. There are 4 possibilities of blood type and persons are selected at random so it is a situation of multinomial distribution. In usual notations, we are given

$p_1 = 0.45,\ p_2 = 0.4,\ p_3 = 0.1,\ p_4 = 0.05;\ x_1 = 25,\ x_2 = 20,\ x_3 = 4,\ x_4 = 1$ and $x_1 + x_2 + x_3 + x_4 = n = 50.$ So, required probability is given by

$$\mathcal{P}\left(X_1 = 25,\ X_2 = 20,\ X_3 = 4,\ X_4 = 1\right)$$

$$= \frac{\lfloor 50}{\lfloor 25 \lfloor 20 \lfloor 4 \lfloor 1}(0.45)^3 (0.4)^2 (0.1)^1 (0.05)^4 = 0.003949808$$

UNIT 11

POISSON AND HYPERGEOMETRIC DISTRIBUTIONS

Structure

11.1 INTRODUCTION

In Unit 10, you studied binomial distribution B(n, p) having two parameters n and p. In this unit, we will see that when n is large, p is small such that np is a fixed finite number then binomial distribution can be approximated by a probability distribution known as Poisson distribution. In Sec. 11.2, we will discuss its PMF and CDF while in Sec. 11.3, we will discuss MGF and some other summary measures of this probability distribution.

In Secs. 11.4 and 11.5, we will do similar studies about hypergeometric distribution. Some applications of these distributions are discussed in Sec. 11.6.

What we have discussed in this unit is summarised in Sec. 11.7. Self-Assessment Questions (SAQs) have been given in some sections which are generally based on the content discussed in that section. But to give you a good practice of what we have discussed in this unit some more questions based on the entire unit are given in Sec. 11.8 under the heading Terminal Questions. Due to the reason mentioned in Sec. 1.1 of Unit 1 of this course, solutions of all the SAQs and Terminal Questions are given in Sec. 11.9.

In the next unit, you will do a similar study about two more discrete probability distributions known as geometric and negative binomial distributions.

Expected Learning Outcomes

After completing this unit, you should be able to:

- ❖ identify situations where you can apply Poisson and hypergeometric distributions;
- ❖ define PMF, CDF and some summary measures of Poisson and hypergeometric distributions; and
- ❖ apply Poisson and hypergeometric distributions to solve problems based on these two probability distributions.

11.2 STORY, DEFINITION, PMF AND CDF OF POISSON DISTRIBUTION

In Sec. 10.2 of the previous unit, you have studied binomial distribution. Recall that in a binomial distribution, we have

- The number of trials ‘n’ is finite in numbers. … (11.1)
- All the ‘n’ trials are independent. … (11.2)
- Probability of success p remains constant in each trial. That is, p does not change from trial to trial. … (11.3)

Now, let us consider one special case where p is small but n is large so that their product np remains a fixed finite number. Let us visualise this idea for different values of n, p and np. Let us first fix the value of np as 1, 2, 3, 4 and 5. Our strategy will be like this for each fix value of np, we will vary n and p and obtain corresponding probabilities using R and then plot PMF of the corresponding binomial distributions. So, the following five cases arise.

Case I: When np = 1. Keeping the condition np = 1 in view let us consider six pair of values of n and p as follows.

(i) n = 10, p = 0.1 (ii) n = 50, p = 0.02 (iii) n = 100, p = 0.01

(iv) n = 1000, p = 0.001 (v) n = 10000, p = 0.0001 (vi) n = 100000, p = 0.00001

Probabilities $\mathcal{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$, if $x = 0, 1, 2, 3, 4, \ldots, n$ using R in each of these six pairs of values of n and p are given as follows.

```
> dbinom(x = 0:10, size = 10, prob = 0.1)
 [1] 0.3486784401 0.3874204890 0.1937102445 0.0573956280 0.0111602610
 [6] 0.0014880348 0.0001377810 0.0000087480 0.0000003645 0.0000000090
[11] 0.0000000001
```

```
> dbinom(x = 0:10, size = 50, prob = 0.02)
 [1] 3.641697e-01 3.716017e-01 1.858009e-01 6.066967e-02 1.454834e-02
 [6] 2.731525e-03 4.180905e-04 5.363260e-05 5.883168e-06 5.603018e-07
[11] 4.688239e-08
```

Here n = 50, so, we could obtain probabilities for values of X from 0 to 50 but note that the probability at x = 10 is 0.00000004688239 which is negligible compared to probabilities at x = 0 or 1. For higher values of X, this probability will be even less than this. That is why, we have obtained only probabilities for X = 0 to 10. Due to the same reason, we are doing so for other pairs of values of n and p. … (11.4)

```
> dbinom(x = 0:10, size = 100, prob = 0.01)
 [1] 3.660323e-01 3.697296e-01 1.848648e-01 6.099917e-02 1.494171e-02
 [6] 2.897787e-03 4.634508e-04 6.286346e-05 7.381694e-06 7.621951e-07
[11] 7.006036e-08

> dbinom(x = 0:10, size = 1000, prob = 0.001)
 [1] 3.676954e-01 3.680635e-01 1.840317e-01 6.128251e-02 1.528996e-02
 [6] 3.048808e-03 5.061001e-04 7.193815e-05 8.938261e-06 9.861812e-07
[11] 9.782838e-08

> dbinom(x = 0:10, size = 10000, prob = 0.0001)
 [1] 3.678610e-01 3.678978e-01 1.839489e-01 6.131017e-02 1.532448e-02
 [6] 3.063976e-03 5.104584e-04 7.288616e-05 9.105303e-06 1.010992e-06
[11] 1.010183e-07

> dbinom(x = 0:10, size = 100000, prob = 0.00001)
 [1] 3.678776e-01 3.678813e-01 1.839406e-01 6.131293e-02 1.532793e-02
 [6] 3.065493e-03 5.108951e-04 7.298137e-05 9.122124e-06 1.013498e-06
[11] 1.013417e-07
```

PMF of these six binomial distributions are shown in Fig. 11.1 (a) to (f).

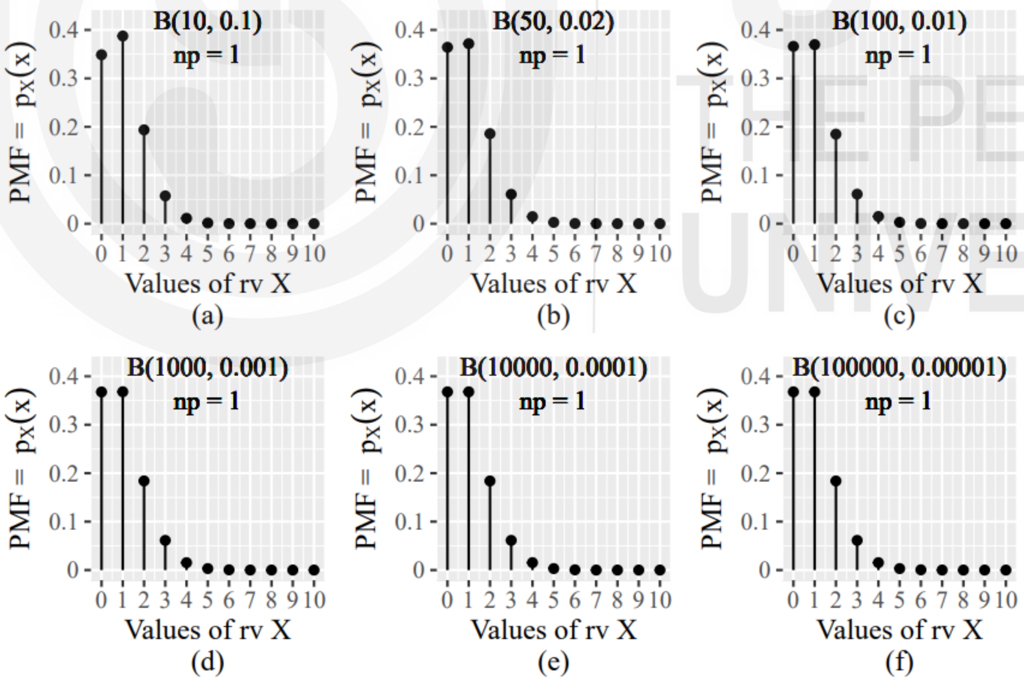


Fig. 11.1: Visualisation of PMF of binomial distributions where np = 1 but n and p vary (a) n = 10, p = 0.1 (b) n = 50, p = 0.02 (c) n = 100, p = 0.01 (d) n = 1000, p = 0.001 (e) n = 10000, p = 0.0001 (f) n = 100000, p = 0.00001

Cases II to V: In cases II to V, we will not obtain probabilities using R as we did in case I. We will directly plot PMF's of binomial distributions where np = 2, 3, 4 and 5 in Figs. 11.2 to 11.5 respectively. The silent features of these PMF's and why we are plotting these PMF's will be discussed after Fig. 11.5 (a) to (f).

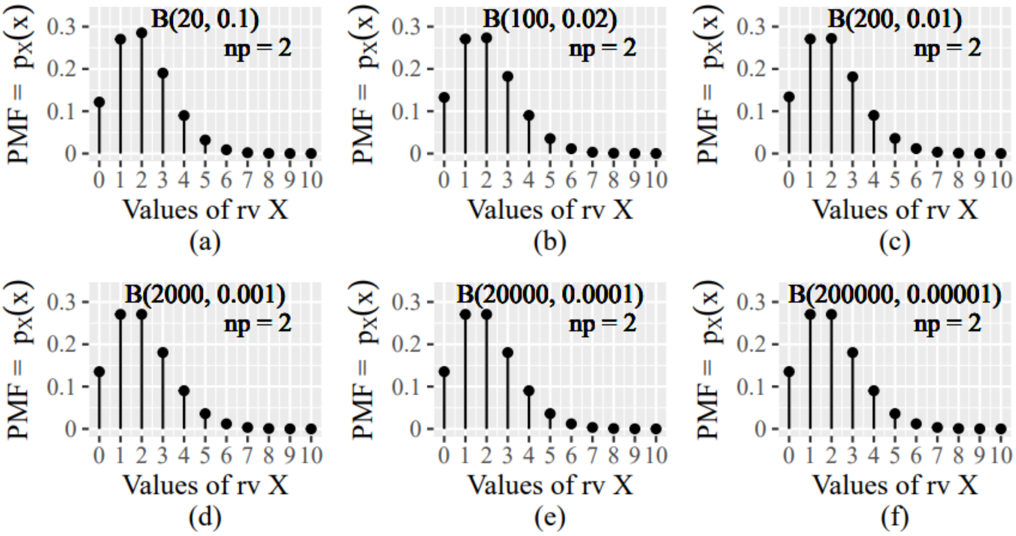


Fig. 11.2: Visualisation of PMF of binomial distributions where np = 2 but n and p vary (a) n = 20, p = 0.1 (b) n = 100, p = 0.02 (c) n = 200, p = 0.01 (d) n = 2000, p = 0.001 (e) n = 20000, p = 0.0001 (f) n = 200000, p = 0.00001

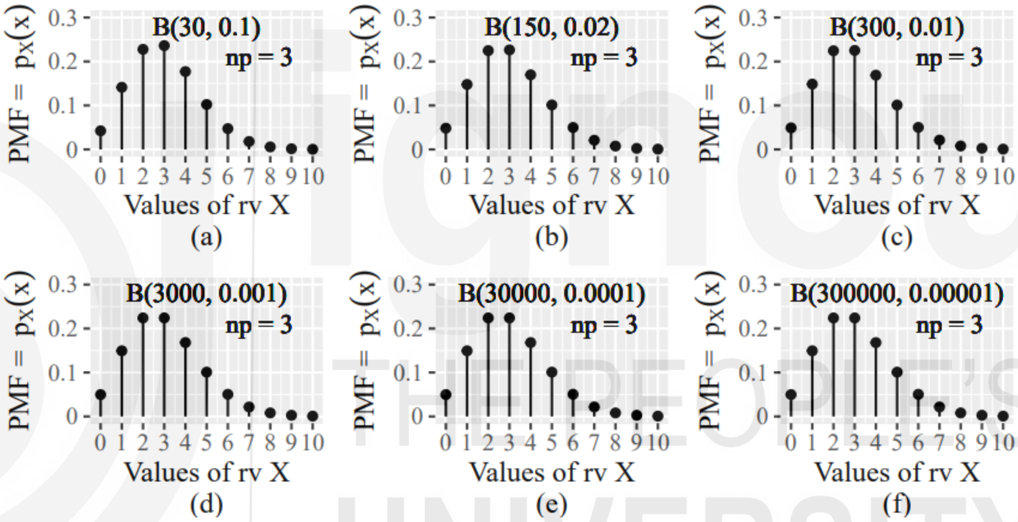


Fig. 11.3: Visualisation of PMF of binomial distributions where np = 3 but n and p vary (a) n = 30, p = 0.1 (b) n = 150, p = 0.02 (c) n = 300, p = 0.01 (d) n = 3000, p = 0.001 (e) n = 30000, p = 0.0001 (f) n = 300000, p = 0.00001

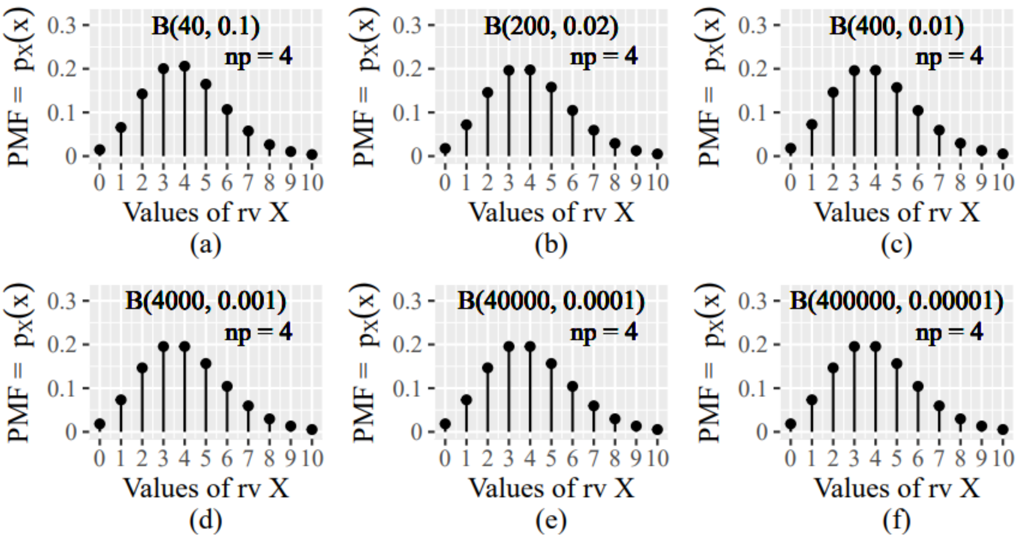


Fig. 11.4: Visualisation of PMF of binomial distributions where np = 4 but n and p vary (a) n = 40, p = 0.1 (b) n = 200, p = 0.02 (c) n = 400, p = 0.001 (d) n = 4000, p = 0.001 (e) n = 40000, p = 0.0001 (f) n = 400000, p = 0.00001

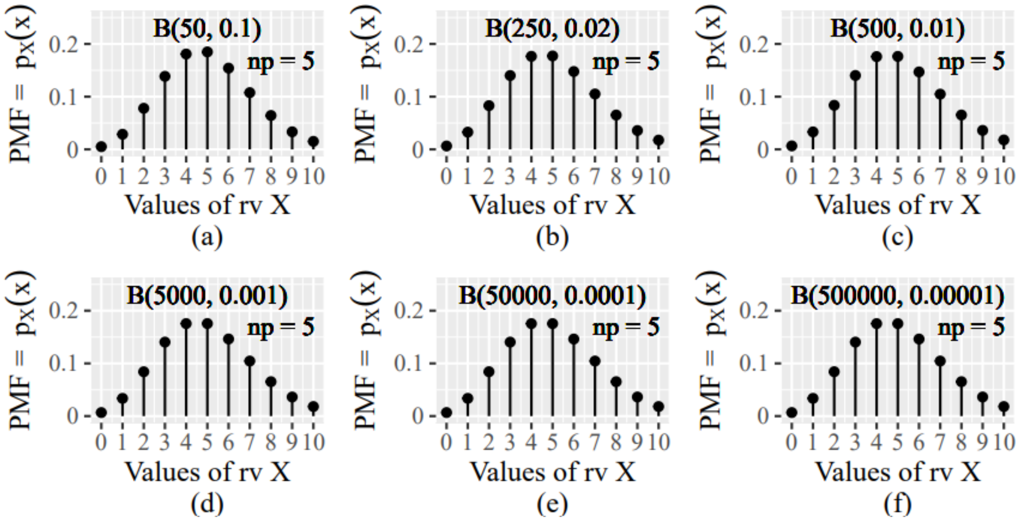


Fig. 11.5: Visualisation of PMF of binomial distributions where np = 5 but n and p vary (a) n = 50, p = 0.1 (b) n = 250, p = 0.02 (c) n = 500, p = 0.01 (d) n = 5000, p = 0.001 (e) n = 50000, p = 0.0001 (f) n = 500000, p = 0.00001

Now, note two important things from these PMF's mentioned as follows.

(a) As $n \geq 50$ and $p < 0.1$ such that np = a fixed number, then for each fixed value of np, the shape of the PMF almost remains the same. You may refer to Figs. 11.1 (b)-(f) to Fig. 11.5 (b)-(f). … (11.5)

(b) As the value of np increases then the symmetry of binomial distribution also increases, you can observe it in Figs. 11.1 to 11.5 for np = 1 to 5 respectively. … (11.6)

For this unit (11.5) is of interest while (11.6) will be used in Unit 14. In view of (11.5), we can say that binomial distribution converges to a particular distribution for each fixed value of np as n approaches to infinity and p is small. This particular distribution is known as the Poisson distribution. Let us obtain the PMF of Poisson distribution from the PMF of binomial distribution under these conditions specified as follows.

$$n \to \infty \qquad \dots (11.7)$$

$$np = \lambda \text{ (a fixed number)} \Rightarrow p = \frac{\lambda}{n} \qquad \dots (11.8)$$

$$p \text{ is small} \qquad \dots (11.9)$$

Using PMF of binomial distribution probability of k successes is given by

$$\mathcal{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ if } k = 0, 1, 2, 3, 4, \dots, n$$

$$= \frac{\underline{|n}}{\underline{|k}\,\underline{|n-k}} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \qquad [\text{Using } (11.8)]$$

$$= \frac{\lambda^k}{\underline{|k}} \frac{n(n-1)(n-2)(n-3)\dots\{n-(k-1)\}\underline{|n-k}}{\underline{|n-k}} \frac{1}{n^k} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= \frac{\lambda^k}{\underline{|k}} n^k \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)\dots\left(1-\frac{k-1}{n}\right)\frac{1}{n^k}\left(\left(1+\frac{-\lambda}{n}\right)^{\frac{n}{-\lambda}}\right)^{-\lambda}\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= \frac{\lambda^k}{\underline{|k}} \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)\dots\left(1-\frac{k-1}{n}\right)\left(\left(1+\frac{-\lambda}{n}\right)^{\frac{n}{-\lambda}}\right)^{-\lambda}\left(1-\frac{\lambda}{n}\right)^{-k}$$

Now, using (11.7), we have

$$\lim_{n\to\infty}\mathcal{P}(X=k)=\lim_{n\to\infty}\frac{\lambda^k}{\underline{|k}}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\left(\left(1+\frac{-\lambda}{n}\right)^{\frac{n}{-\lambda}}\right)^{-\lambda}\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$=\frac{\lambda^k}{\underline{|k}}\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\left(\lim_{n\to\infty}\left(1+\frac{-\lambda}{n}\right)^{\frac{n}{-\lambda}}\right)^{-\lambda}\left(\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)\right)^{-k}$$

$$=\frac{\lambda^k}{\underline{|k}}(1-0)(1-0)(1-0)\ldots(1-0)(e)^{-\lambda}(1-0)^{-k}\quad\left[\because \lim_{x\to\infty}\left(1+\frac{1}{x}\right)^x=e\right]$$

$$=\frac{e^{-\lambda}\lambda^k}{\underline{|k}},\ k=0,1,2,3,\ldots$$

Hence, PMF of binomial distribution under the conditions (11.7) to (11.9) converges to the PMF

$$p_X(x)=\mathcal{P}(X=k)=\frac{e^{-\lambda}\lambda^k}{\underline{|k}},\ k=0,1,2,3,\ldots \qquad \ldots (11.10)$$

The PMF given by (11.10) as already mentioned is known as PMF of Poisson distribution.

Now, we can discuss the story of the Poisson distribution.

Story of Poisson Distribution: If we have a large number of Bernoulli trials with small probability success and np = a fixed finite number then corresponding binomial distribution can be approximated by a probability distribution known as Poisson distribution.

Definition and PMF of Poisson Distribution: If X ~ B(n, p) such that n approaches to infinity, p is small and $np=\lambda$ is a fixed finite number then binomial PMF converges to the PMF given by

$$p_X(x)=\mathcal{P}(X=k)=\begin{cases}\dfrac{e^{-\lambda}\lambda^k}{\underline{|k}}, & k=0,1,2,3,\ldots\\ 0, & \text{otherwise}\end{cases} \qquad \ldots (11.11)$$

The PMF given by (11.11) is known as PMF of Poisson distribution with parameter λ and is denoted by writing $X\sim\text{Pois}(\lambda)$. … (11.12)

Like the Bernoulli distribution case, we read $X\sim\text{Pois}(\lambda)$ as X is distributed as a Poisson distribution with parameter $\lambda.$ Or we read it as X follows a Poisson distribution with parameter $\lambda.$ … (11.13)

Since the statistical software used for hands-on practice for the lab courses of this programme is R programming language, so, you should also know its notation in R. In R notation that is used for Poisson distribution is pois(lambda) in the stats package, where lambda represents the value of mean or rate of the Poisson distribution. In fact, there are four functions for Poisson distribution namely dpois(x, lambda, …), ppois(q, lambda, …), qpois(p, lambda, …), and rpois(n, lambda, …). We have already explained the meaning of these functions in Unit 9. … (11.14)

Let us check the **validity of the PMF of the Poisson distribution**.

We have to check two conditions non-negativity and normality refer to (4.62).

(1) **Non-negativity**: Since exponential and factorial functions take non-negative values hence

$$\frac{e^{-\lambda}\lambda^k}{\lfloor k} \ge 0,\ k = 0, 1, 2, 3, \ldots \qquad \ldots (11.15)$$

(2) **Normality**: Let us obtain sum of all probabilities of Poisson distribution.

$$\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{\lfloor k} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{\lfloor k} = e^{-\lambda}\left(\frac{\lambda^0}{\lfloor 0}+\frac{\lambda^1}{\lfloor 1}+\frac{\lambda^2}{\lfloor 2}+\frac{\lambda^3}{\lfloor 3}+\ldots\right) = e^{-\lambda}\left(1+\frac{\lambda}{\lfloor 1}+\frac{\lambda^2}{\lfloor 2}+\frac{\lambda^3}{\lfloor 3}+\ldots\right)$$

$$= e^{-\lambda}e^{\lambda}\left[\text{Refer (1.31) of MST-011, } e^x = 1 + x + \frac{x^2}{\lfloor 2}+\ldots\right]$$

$$= e^{-\lambda+\lambda} = e^0 = 1 \qquad \ldots (11.16)$$

This proves that sum of all probabilities of Poisson distribution is 1. So, we can say that PMF of the Poisson random variable is a valid PMF.

Now, we define the CDF of Poisson distribution.

CDF of Poisson Distribution: If X ~ B(n, p) such that n approaches to infinity, p is small and $np = \lambda$ is a fixed finite number then binomial PMF converges to the PMF given by

$$p_X(x) = \mathcal{P}(X = k) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{\lfloor k}, & k = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

This PMF is known as PMF of Poisson distribution with parameter λ.

Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}(X \le x) = \begin{cases} 0, & \text{if } x < 0 \\ \displaystyle\sum_{k=0}^{[x]}\frac{e^{-\lambda}\lambda^k}{\lfloor k}, & \text{if } 0 \le x < \infty \\ 1, & \text{if } x \to \infty \end{cases} \qquad \ldots (11.17)$$

Let us do one example.

Example 1: In a company there are 600 employees. On the basis of the past experience, it is known that an employee of this company will be absent on any one day is 0.01. Find the probability that the number of employees that are absent on any one day is (i) 0 (ii) 1 (iii) 2 (iv) 3 (v) 4 (vi) 5 (vii) 6 (viii) 7 (ix) 8 (x) 9 (xi) 10 (xii) 11.

Solution: If we define getting an employee of this company absent as success then probability of success is p = 0.01. The number of employees in this company is n = 600. Also, np = 600(0.01) = 6 = λ will remain the same. So, here n is large, p is small and np = a fixed finite number. So, binomial distribution B(600, 0.01) can be approximated by Pois(6 = np = λ). So, using the PMF of Poisson distribution required probabilities can be obtained as follows.

$$\mathcal{P}(X = 0) = \frac{e^{-6}6^0}{\lfloor 0} = 0.002478752 \qquad \mathcal{P}(X = 6) = \frac{e^{-6}6^6}{\lfloor 6} = 0.16062314$$

$$\mathcal{P}(X = 1) = \frac{e^{-6}6^1}{\lfloor 1} = 0.01487251 \qquad \mathcal{P}(X = 7) = \frac{e^{-6}6^7}{\lfloor 7} = 0.13767698$$

$$\mathcal{P}(X=2)=\frac{e^{-6}6^{2}}{\lfloor 2}=0.04461754 \qquad \mathcal{P}(X=8)=\frac{e^{-6}6^{8}}{\lfloor 8}=0.10325773$$

$$\mathcal{P}(X=3)=\frac{e^{-6}6^{3}}{\lfloor 3}=0.08923508 \qquad \mathcal{P}(X=9)=\frac{e^{-6}6^{9}}{\lfloor 9}=0.06883849$$

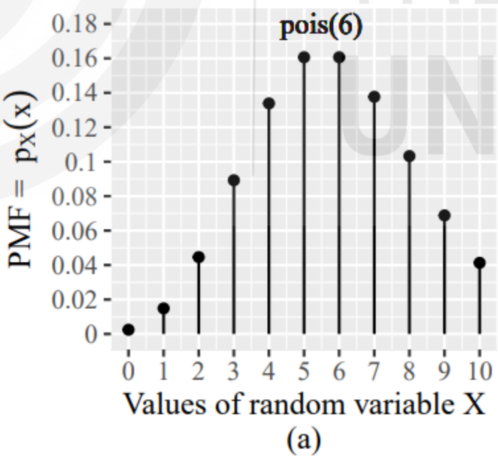
$$\mathcal{P}(X=4)=\frac{e^{-6}6^{4}}{\lfloor 4}=0.13385262 \qquad \mathcal{P}(X=10)=\frac{e^{-6}6^{10}}{\lfloor 10}=0.04130309$$

$$\mathcal{P}(X=5)=\frac{e^{-6}6^{5}}{\lfloor 5}=0.16062314 \qquad \mathcal{P}(X=11)=\frac{e^{-6}6^{11}}{\lfloor 11}=0.02252896$$

Now, using (11.17) CDF of X is given by

$$F_X(x)=\mathcal{P}(X\le x)=\begin{cases}0, & \text{if } x<0\\ 0.00247875, & \text{if } 0\le x<1\\ 0.01735127, & \text{if } 1\le x<2\\ 0.06196880, & \text{if } 2\le x<3\\ 0.15120388, & \text{if } 3\le x<4\\ 0.28505650, & \text{if } 4\le x<5\\ 0.44567964, & \text{if } 5\le x<6\\ 0.60630278, & \text{if } 6\le x<7\\ 0.74397976, & \text{if } 7\le x<8\\ 0.84723749, & \text{if } 8\le x<9\\ 0.91607598, & \text{if } 9\le x<10\\ 0.95737908, & \text{if } 10\le x<11\\ 0.97990804, & \text{if } 11\le x<12\\ \vdots & \end{cases} \quad \ldots (11.18)$$

PMF and CDF are plotted in Fig. 11.6 (a) and (b) respectively as follows.



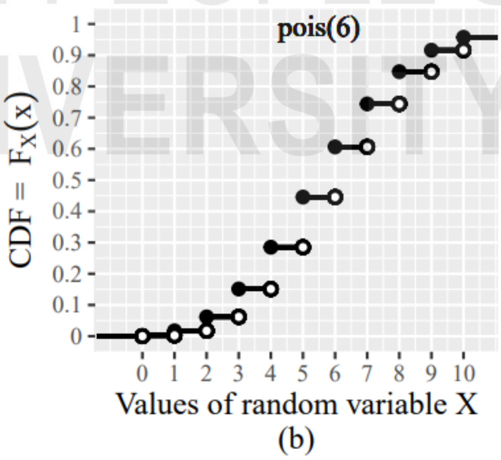


Fig. 11.6: Visualisation of (a) PMF (b) CDF of the Pois(6) discussed in Example 1

11.3 MGF AND OTHER SUMMARY MEASURES OF POISSON DISTRIBUTION

In the previous section, you have studied PMF and CDF of Poisson distribution. In this section, we want to obtain MGF and some other summary measure of Poisson distribution like mean and variance. Let us first obtain MGF of Poisson distribution. We will obtain MGF of Poisson distribution using definition of MGF refer to (7.48).

Calculation of MGF

$$M_X(t) = E\left(e^{tX}\right),\ t \in \mathbb{R}$$

$$= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{\lfloor x} = e^{-\lambda}\sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x}{\lfloor x} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\left(\lambda e^t\right)^x}{\lfloor x}$$

$$= e^{-\lambda}\left(\frac{\left(\lambda e^t\right)^0}{\lfloor 0} + \frac{\left(\lambda e^t\right)^1}{\lfloor 1} + \frac{\left(\lambda e^t\right)^2}{\lfloor 2} + \frac{\left(\lambda e^t\right)^3}{\lfloor 3} + \dots\right)$$

$$= e^{-\lambda}\left(e^{\lambda e^t}\right) \quad \left[\text{Refer (1.31) of MST-011, } e^x = 1 + x + \frac{x^2}{\lfloor 2} + \frac{x^3}{\lfloor 3} + \dots\right]$$

$$= e^{\lambda\left(e^t - 1\right)}$$

$$M_X(t) = e^{\lambda\left(e^t - 1\right)} \quad \dots (11.19)$$

Calculation for Mean

$$\text{Expected value} = E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{\lfloor x} = \sum_{x=1}^{\infty} x\frac{e^{-\lambda}\lambda^x}{\lfloor x} \left[\because \text{ when } x = 0, \text{ then we get value of } x\frac{e^{-\lambda}\lambda^x}{\lfloor x} = 0\right]$$

$$= e^{-\lambda}\sum_{x=1}^{\infty} x \frac{\lambda\,\lambda^{x-1}}{x\lfloor x-1} = \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{\lfloor x-1}$$

$$= \lambda e^{-\lambda}\left(\frac{(\lambda)^0}{\lfloor 0} + \frac{(\lambda)^1}{\lfloor 1} + \frac{(\lambda)^2}{\lfloor 2} + \frac{(\lambda)^3}{\lfloor 3} + \dots\right) = \lambda e^{-\lambda}\left(1 + \lambda + \frac{(\lambda)^2}{\lfloor 2} + \frac{(\lambda)^3}{\lfloor 3} + \dots\right)$$

$$= \lambda e^{-\lambda}\left(e^{\lambda}\right) \quad \left[\text{Refer (1.31) of MST-011, } e^x = 1 + x + \frac{x^2}{\lfloor 2} + \frac{x^3}{\lfloor 3} + \dots\right]$$

$$= \lambda e^{-\lambda+\lambda} = \lambda e^0 = \lambda(1) = \lambda$$

Hence, mean of X = expected value of $X = E(X) = \lambda$. … (11.20)

Calculation for Variance and Standard Deviation

$$E\left(X(X-1)\right) = \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{\lfloor x} = \sum_{x=2}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{\lfloor x} \left[\because \text{ when } x = 0, 1 \text{ then value of } x(x-1)\frac{e^{-\lambda}\lambda^x}{\lfloor x} = 0\right]$$

$$= e^{-\lambda}\sum_{x=2}^{\infty} x(x-1)\frac{\lambda^2\lambda^{x-2}}{x(x-1)\lfloor x-2} = \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{\lfloor x-2}$$

$$= \lambda^2 e^{-\lambda}\left(\frac{(\lambda)^0}{\lfloor 0} + \frac{(\lambda)^1}{\lfloor 1} + \frac{(\lambda)^2}{\lfloor 2} + \frac{(\lambda)^3}{\lfloor 3} + \dots\right) = \lambda^2 e^{-\lambda}\left(1 + \lambda + \frac{(\lambda)^2}{\lfloor 2} + \frac{(\lambda)^3}{\lfloor 3} + \dots\right)$$

$$= \lambda^2 e^{-\lambda}\left(e^{\lambda}\right) \quad \left[\text{Refer (1.31) of MST-011, } e^x = 1 + x + \frac{x^2}{\lfloor 2} + \frac{x^3}{\lfloor 3} + \dots\right]$$

$$= \lambda^2 e^{-\lambda+\lambda} = \lambda^2 e^0 = \lambda^2(1) = \lambda^2 \quad \dots (11.21)$$

$$E\left(X^2\right) = E\left(X(X-1) + X\right) = E\left(X(X-1)\right) + E(X) \quad \left[\text{Using addition theorem of expectation refer to (7.29)}\right]$$

$$= \lambda^2 + \lambda \quad \left[\text{Using (11.20) and (11.21)}\right]$$

$$\Rightarrow E\left(X^2\right) = \lambda^2 + \lambda \qquad \ldots (11.22)$$

Using (7.63) variance of any random variable X is given by

$$\text{Variance of the Poisson distribution} = \mu_2 = E\left(X^2\right) - \left(E(X)\right)^2$$
$$= \lambda^2 + \lambda - (\lambda)^2 \qquad [\text{Using (11.22) and (11.20)}]$$
$$= \lambda^2 + \lambda - \lambda^2 = \lambda$$

Hence, variance of the Poisson distribution $= \lambda$ … (11.23)

We know that standard deviation of X is positive square root of variance of X.
Hence, $SD(X) = \sqrt{\text{Variance of } X} = \sqrt{\lambda}$ … (11.24)

Similarly, we can obtain other measures of our interest. But keeping applied nature of the programme in view, we are not focusing on proof of each measure. Some commonly used summary measures of Poisson distribution are shown in Table 11.1 given as follows.

Table 11.1: Summary measures of Poisson distribution

Name of measure	**Formula**	**Name of measure**	**Formula**
Mean	λ	MGF	$e^{\lambda\left(e^t - 1\right)}$
Variance	λ	Skewness	$\frac{1}{\sqrt{\lambda}}$
Standard deviation	$\sqrt{\lambda}$	Kurtosis	$3 + \frac{1}{\lambda}$

11.4 STORY, DEFINITION, PMF AND CDF OF HYPERGEOMETRIC DISTRIBUTION

Recall that in binomial distribution, we were using sampling with replacement. With replacement sampling scheme makes probability of each draw independent. For example, suppose in a bag there are 6 red balls and 4 black balls. If R_i and B_i denote the event of getting a red and black ball respectively in i^{th} draw, $i = 1, 2, 3, \ldots$, then using concepts of Unit 1, we have

$$\mathcal{P}\left(R_i\right) = \frac{\binom{6}{1}}{\binom{10}{1}} = \frac{6}{10} = 0.6, \text{ and } \mathcal{P}\left(B_i\right) = \frac{\binom{4}{1}}{\binom{10}{1}} = \frac{4}{10} = 0.4, \; i = 1, 2, 3, \ldots \qquad \ldots (11.25)$$

So, note that probability of getting a red ball under sampling with replacement is 0.6 in each draw and that of black ball is 0.4 in each draw. … (11.26)

But if, we draw using without replacement sampling scheme then using concepts of conditional probability discussed in Unit 1, we have

Probability of getting a red ball in the first draw $= \mathcal{P}\left(R_1\right) = \dfrac{\binom{6}{1}}{\binom{10}{1}} = \dfrac{6}{10} = 0.6.$

… (11.27)

Probability of getting a red ball in the second draw given that first drawn ball

was red = $\mathcal{P}\left(R_2 \mid R_1\right) = \dfrac{\binom{5}{1}}{\binom{9}{1}} = \dfrac{5}{9}.$ … (11.28)

Probability of getting a red ball in the second draw given that first drawn ball

was black = $\mathcal{P}\left(R_2 \mid B_1\right) = \dfrac{\binom{6}{1}}{\binom{9}{1}} = \dfrac{6}{9} = \dfrac{2}{3}.$ … (11.29)

and so on. Similarly, we can obtain probabilities of getting a black ball in different draws. Note that in without replacement sampling scheme probability of getting a red ball in the first draw is different from getting a red ball in the second draw. So, in without replacement sampling scheme probabilities of a ball of a particular colour is different in different draws. It means here probabilities are dependent in different trials. That is probabilities are not independent trial to trial like they were in the case of sampling with replacement scheme and in that case, we were using binomial distribution.

… (11.30)

Now, we can write the story of hypergeometric distribution as follows.

Story of Hypergeometric Distribution: If a bag contains G good items and B bad items and, we are interested in probability of getting g good items out of n drawn items from this bag where sampling is done without replacement then the probability distribution which model this situation is known as hypergeometric distribution.

Now, before defining hypergeometric distribution, first, we have to explain how sampling n (= g + b) items from a population having N (= G + B) items where G items are good and B items are bad one by one without replacement is equivalent to sampling n (= g + b) items all at a time (simultaneously) where g out of n items are good and remaining b = n – g items are bad. Let us first obtain the probability when items are drawn simultaneously. Therefore, probability of getting n items out of N = G + B items where g items are good and remaining b = n – g items are bad is given by

$$\frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}} = \frac{\dfrac{\lfloor G}{\lfloor g \lfloor G-g} \dfrac{\lfloor B}{\lfloor b \lfloor B-b}}{\dfrac{\lfloor N}{\lfloor n \lfloor N-n}} = \frac{\lfloor G \lfloor B \lfloor n \lfloor N-n}{\lfloor g \lfloor b \lfloor G-g \lfloor B-b \lfloor N} = \frac{\lfloor n}{\lfloor g \lfloor b} \times \frac{\lfloor G}{\lfloor G-g} \times \frac{\lfloor B}{\lfloor B-b} \times \frac{\lfloor N-n}{\lfloor N}$$

$$= \frac{\lfloor n}{\lfloor g \lfloor n-g} \times \frac{\lfloor G}{\lfloor G-g} \times \frac{\lfloor B}{\lfloor B-b} \times \frac{\lfloor N-n}{\lfloor N} \qquad \left[\because b = n-g\right]$$

$$= \binom{n}{g} \times \frac{\underline{|G}}{\underline{|G-g}} \times \frac{\underline{|B}}{\underline{|B-b}} \times \frac{\underline{|N-n}}{\underline{|N}} \quad \ldots (11.31)$$

To obtain the expression when we draw items one by one without replacement. Let us consider three particular sequences of getting g good and b bad items one by one without replacement where n = g + b.

Case I: In the first g draws, we get good items and in last b draws, we get bad items where g + b = n. Probability of this particular sequence is given by

$$\frac{G}{N} \times \frac{G-1}{N-1} \times \frac{G-2}{N-2} \times \ldots \times \frac{G-(g-1)}{N-(g-1)} \times \frac{B}{N-g} \times \frac{B-1}{N-g-1} \times \frac{B-2}{N-g-2} \times \ldots \times \frac{B-(b-1)}{N-g-(b-1)} \quad \ldots (11.32)$$

Case II: In the first b draws, we get b bad items and in the last g draws, we get g good items where g + b = n. Probability of this particular sequence is given by

$$\frac{B}{N} \times \frac{B-1}{N-1} \times \frac{B-2}{N-2} \times \ldots \times \frac{B-(b-1)}{N-(b-1)} \times \frac{G}{N-b} \times \frac{G-1}{N-b-1} \times \frac{G-2}{N-b-2} \times \ldots \times \frac{G-(g-1)}{N-b-(g-1)} \quad \ldots (11.33)$$

Case III: First drawn item is good and then from second to $(b+1)^{th}$ all items are continuously bad and finally last g – 1 items are good where g + b = n. Probability of this particular sequence is given by

$$\frac{G}{N} \times \frac{B}{N-1} \times \frac{B-1}{N-2} \times \frac{B-2}{N-3} \times \ldots \times \frac{B-(b-1)}{N-b} \times \frac{G-1}{N-b-1} \times \frac{G-2}{N-b-2} \times \ldots \times \frac{G-(g-1)}{N-b-(g-1)} \quad \ldots (11.34)$$

From (11.32) to (11.34) note that all the three numerators have exactly the same factors except their order of presence. Also, all the three denominators in (11.32) to (11.34) have exactly the same factors except their order of presence. Now it is easy to count how many such sequences with g good draws and b bad draws where g + b = n are possible. This is equivalent to select g positions for good draws out of the total n positions. We know that this can be done in $\binom{n}{g}$ number of ways. Hence, the probability of getting g good items and b bad items, i.e., total n = g + b items out of total N = G + B items one by one without replacement is given by

$$\binom{n}{g} \frac{G(G-1)(G-2)\ldots(G-g+1)B(B-1)(B-2)\ldots(B-b+1)}{N(N-1)(N-2)\ldots(N-g-b+1)}$$

$$= \binom{n}{g} \frac{G(G-1)(G-2)\ldots(G-g+1)}{1} \times \frac{(G-g)(G-g-1)(G-g-2)\ldots3.2.1}{(G-g)(G-g-1)(G-g-2)\ldots3.2.1}$$

$$\times \frac{B(B-1)(B-2)\ldots(B-b+1)}{1} \frac{(B-b)(B-b-1)(B-b-2)\ldots3.2.1}{(B-b)(B-b-1)(B-b-2)\ldots3.2.1}$$

$$\times \frac{(N-n)(N-n-1)(N-n-2)\ldots3.2.1}{N(N-1)(N-2)\ldots(N-n+1)(N-n)\ldots3.2.1} \left[\because g+b=n\right]$$

$$= \binom{n}{g} \times \frac{\underline{|G}}{\underline{|G-g}} \times \frac{\underline{|B}}{\underline{|B-b}} \times \frac{\underline{|N-n}}{\underline{|N}} \quad \ldots (11.35)$$

From (11.31) and (11.35), we can say that sampling n (= g + b) items from a population having N (= G + B) items where G items are good and B items are bad one by one without replacement is equivalent to sampling n (= g + b) items simultaneously where g out of n items are good and remaining b = n – g items are bad.

Now, we define hypergeometric distribution as follows.

Definition and PMF of Hypergeometric Distribution: Suppose in a bag there are total N items out of which G items are good and remaining B = N – G items are bad. We draw n items randomly from the given bag having N items one by one without replacement. If we are interested in the probability of getting g good items out of n = g + b = drawn items where b represents number of bad items then it is given by $\dfrac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}$. … (11.36)

So, PMF, $p_X(x)$ of hypergeometric distribution is given by

$$p_X(x) = \mathcal{P}(X = x) = \begin{cases} \dfrac{\binom{G}{x}\binom{B}{n-x}}{\binom{N}{n}}, & \text{if } x = 0, 1, 2, \ldots, \min\{G, n\} \\ 0, & \text{otherwise} \end{cases} \quad \ldots (11.37)$$

If random variable X has PMF given by (11.37), then we say that it follows hypergeometric distribution with parameters G, B and n and is denoted by writing $X \sim \text{HGeom}(G, B, n)$. … (11.38)

We read it as X follows hypergeometric distribution with parameters G, B and n. … (11.39)

Since the statistical software used for hands-on practice for the lab courses of this programme is R programming language, so, you should also know its notation in R. In R notation that is used for hypergeometric distribution is hyper(m, n, k) in the stats package, where m represents the number of good items, n represents the number of bad items in the bag and k represents size of the sample. In fact, there are four functions for hypergeometric distribution namely dhyper(x, m, n, k, …), phyper(q, m, n, k, …), qhyper(p, m, n, k, …) and rhyper(nn, m, n, k, …). We have already explained meaning of these functions in Unit 9. ... (11.40)

Let us check the **validity of the PMF of the Hypergeometric distribution**.

In checking the validity of PMF of hypergeometric distribution, we will use Vandermonde's Identity. So, let first state and prove it.

Vandermonde's Identity: Let G, B be the number of good and bad items in a bag and, we have drawn n items from the bag then prove that

$$\sum_{g=0}^{n} \binom{G}{g}\binom{B}{n-g} = \binom{G+B}{n}, \quad \ldots (11.41)$$

where obviously G, B, n are positive integers and $n \le G + B$

Proof: The result given by (11.41) is a combinatorics result. So, we can prove it in two ways (a) Algebraic proof (b) Story proof.

Let us follow second approach. So, we have to tell two stories one for LHS and one for RHS such that both counts the same thing but in two different ways. That is counting of one story should match RHS of (11.41) and counting of another story should match with LHS of (11.41).

Story 1 for RHS: If in a bag there are total G + B number of items then the number of ways of selecting n of them will be $\binom{G+B}{n}$ ways. … (11.42)

Story 2 for LHS: Out of the selected n items the number of good items may be 0, 1, 2, 3, …, n. If good items are 0 then obviously the number of bad items will be n. Similarly, if good item is 1 then obviously the number of bad items will be n – 1. If good items are 2 then obviously the number of bad items will be n – 2, and so on if good items are n then obviously the number of bad items will be 0. So, using fundamental principle of multiplication and addition total number of ways of selecting n items from the bag are $\sum_{g=0}^{n}\binom{G}{g}\binom{B}{n-g}$. … (11.43)

Since (11.42) and (11.43) both count the same thing and hence

$\sum_{g=0}^{n}\binom{G}{g}\binom{B}{n-g}=\binom{G+B}{n}$.

Now, we can check the **validity of the PMF of the Hypergeometric distribution**.

We have to check two conditions non-negativity and normality refer to (4.62).

(1) Non-negativity: Since

$$\binom{G}{g}>0,\ \binom{B}{b}>0 \text{ and } \binom{N}{n}>0, \text{ so } \frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}>0\ \ \forall\ g=0,1,2,\ldots,\min\{G,n\} \quad \ldots (11.44)$$

(2) **Normality**: Using (11.41), we have

$$\sum_{g=0}^{n}\binom{G}{g}\binom{B}{n-g}=\binom{G+B}{n} \quad \ldots (11.45)$$

This proves that $\dfrac{\sum_{g=0}^{n}\binom{G}{g}\binom{B}{n-g}}{\binom{G+B}{n}}=1.$ Hence, sum of all probabilities of hypergeometric distribution is 1.

Hence, PMF defined by (11.37) of hypergeometric distribution is a valid PMF.

Now, we define CDF of hypergeometric distribution.

CDF of Hypergeometric Distribution: If $X \sim \text{HGeom}(G, B, n)$, then PMF of X is given by

$$p_X(x) = \mathcal{P}(X = x) = \begin{cases} \dfrac{\dbinom{G}{x}\dbinom{B}{n-x}}{\dbinom{N}{n}}, & \text{if } x = 0, 1, 2, \ldots, \min\{G, n\} \\ 0, & \text{otherwise} \end{cases}$$

Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}(X \le x) = \begin{cases} 0, & \text{if } x < 0 \\ \dfrac{\sum\limits_{k=0}^{[x]}\dbinom{G}{k}\dbinom{B}{n-k}}{\dbinom{N}{n}}, & \text{if } 0 \le x < \min\{G, n\} \\ 1, & \text{if } x \ge \min\{G, n\} \end{cases} \qquad \ldots (11.46)$$

11.5 MGF AND OTHER SUMMARY MEASURES OF HYPERGEOMETRIC DISTRIBUTION

In the previous section, you have studied PMF and CDF of hypergeometric distribution. In this section, we want to obtain mean and variance of hypergeometric distribution.

Calculation of Mean

By definition of expected value, we have

$$E(X) = \sum_{x=0}^{n} x \frac{\dbinom{G}{x}\dbinom{B}{n-x}}{\dbinom{G+B}{n}} = \frac{1}{\dbinom{G+B}{n}} \sum_{x=0}^{n} x \binom{G}{x}\binom{B}{n-x} \quad \left[\because \binom{G+B}{n} \text{ is free from } x\right]$$

$$= \frac{1}{\dbinom{G+B}{n}} \sum_{x=0}^{n} x \frac{G}{x}\binom{G-1}{x-1}\binom{B}{n-x} \qquad \left[\because \binom{G}{x} = \frac{G}{x}\binom{G-1}{x-1}\right]$$

$$= \frac{G}{\dbinom{G+B}{n}} \sum_{x=1}^{n} \binom{G-1}{x-1}\binom{B}{n-x} \qquad \left[\because \text{when } x = 0, \text{ then } \binom{G-1}{x-1} = 0\right]$$

$$= \frac{G}{\dbinom{G+B}{n}} \binom{G+B-1}{n-1} \qquad [\text{Using Vandermonde's identity}]$$

$$= \frac{G}{\dfrac{\underline{|G+B}}{\underline{|n}\,\underline{|G+B-n}}} \times \frac{\underline{|G+B-1}}{\underline{|n-1}\,\underline{|G+B-1-(n-1)}} = \frac{nG\,\underline{|n-1}\,\underline{|G+B-n}}{(G+B)\underline{|G+B-1}} \times \frac{\underline{|G+B-1}}{\underline{|n-1}\,\underline{|G+B-n}}$$

$$= \frac{nG}{(G+B)}$$

$$\text{Hence, mean of } X = \text{expected value of } X = E(X) = \frac{nG}{(G+B)} \qquad \ldots (11.47)$$

Calculation of Variance

$$E\left(X(X-1)\right)=\sum_{x=0}^{n}x(x-1)\frac{\binom{G}{x}\binom{B}{n-x}}{\binom{G+B}{n}}=\frac{1}{\binom{G+B}{n}}\sum_{x=0}^{n}x(x-1)\binom{G}{x}\binom{B}{n-x}$$

$$=\frac{1}{\binom{G+B}{n}}\sum_{x=0}^{n}x(x-1)\frac{G(G-1)}{x(x-1)}\binom{G-2}{x-2}\binom{B}{n-x}\left[\because \binom{G}{x}=\frac{G(G-1)}{x(x-1)}\binom{G-2}{x-2}\right]$$

$$=\frac{G(G-1)}{\binom{G+B}{n}}\sum_{x=2}^{n}\binom{G-2}{x-2}\binom{B}{n-x}\qquad\left[\because \text{ when } x=0,1 \text{ then } \binom{G-2}{x-2}=0\right]$$

$$=\frac{G(G-1)}{\binom{G+B}{n}}\binom{G+B-2}{n-2}\qquad\left[\text{Using Vandermonde's identity}\right]$$

$$=\frac{G(G-1)}{\dfrac{\lfloor G+B}{\lfloor n\lfloor G+B-n}}\times\frac{\lfloor G+B-2}{\lfloor n-2\lfloor G+B-2-(n-2)}$$

$$=\frac{G(G-1)n(n-1)\lfloor n-2\lfloor G+B-n}{(G+B)(G+B-1)\lfloor G+B-2}\times\frac{\lfloor G+B-2}{\lfloor n-2\lfloor G+B-n}$$

$$=\frac{n(n-1)G(G-1)}{(G+B)(G+B-1)}\qquad\ldots(11.48)$$

$$E\left(X^2\right)=E\left(X(X-1)+X\right)=E\left(X(X-1)\right)+E\left(X\right)\left[\begin{array}{l}\text{Using addition theorem of}\\ \text{expectation refer to (7.29)}\end{array}\right]$$

$$=\frac{n(n-1)G(G-1)}{(G+B)(G+B-1)}+\frac{nG}{(G+B)}\left[\text{Using (11.47) and (11.48)}\right]\qquad\ldots(11.49)$$

Now,

$$V(X)=E\left(X^2\right)-\left(E(X)\right)^2$$

$$=\frac{n(n-1)G(G-1)}{(G+B)(G+B-1)}+\frac{nG}{(G+B)}-\left(\frac{nG}{(G+B)}\right)^2\left[\begin{array}{l}\text{Using (11.49)}\\ \text{and (11.47)}\end{array}\right]$$

$$=\frac{n(n-1)G(G-1)(G+B)+nG(G+B)(G+B-1)-n^2G^2(G+B-1)}{(G+B)^2(G+B-1)}$$

$$=\frac{nG(G+B)\left\{(n-1)(G-1)+G+B-1\right\}-n^2G^2(G+B)+n^2G^2}{(G+B)^2(G+B-1)}$$

$$=\frac{nG(G+B)\left\{nG-n+B\right\}-n^2G^2(G+B)+n^2G^2}{(G+B)^2(G+B-1)}$$

$$= \frac{n^2G^2(G+B) + nG(G+B)(-n+B) - n^2G^2(G+B) + n^2G^2}{(G+B)^2(G+B-1)}$$

$$= \frac{nG(-nG+GB-nB+B^2) + n^2G^2}{(G+B)^2(G+B-1)}$$

$$= \frac{-n^2G^2 + nG(GB-nB+B^2) + n^2G^2}{(G+B)^2(G+B-1)} = \frac{nGB(G-n+B)}{(G+B)^2(G+B-1)}$$

$$= \frac{nGB(G+B-n)}{(G+B)^2(G+B-1)}$$

Hence, variance of $X = V(X) = \dfrac{nGB(G+B-n)}{(G+B)^2(G+B-1)}$ … (11.50)

We know that standard deviation of X is positive square root of variance of X.

Hence, standard deviation $= SD(X) = \sqrt{\text{Variance of } X} = \sqrt{\dfrac{nGB(G+B-n)}{(G+B)^2(G+B-1)}}$

… (11.51)

Let us put these calculated summary measures of hypergeometric distribution in Table 11.2 given as follows.

Table 11.2: Summary measures of hypergeometric distribution

Name of measure	**Formula**	**Name of measure**	**Formula**
Mean	$\dfrac{nG}{(G+B)}$	Standard deviation	$\sqrt{\dfrac{nGB(G+B-n)}{(G+B)^2(G+B-1)}}$
Variance	$\dfrac{nGB(G+B-n)}{(G+B)^2(G+B-1)}$		

11.6 APPLICATIONS OF POISSON AND HYPERGEOMETRIC DISTRIBUTIONS

In this section, we will apply Poisson and hypergeometric distributions to solve some problems where assumptions of these distributions make sense.

Example 2: **Application in solving problems related to playing cards**: From a well shuffled pack of 52 playing cards, 39 cards are drawn one by one without replacement. Find the expected number of jack or queen or king out of these 39 selected cards.

Solution: Here sampling is done without replacement and cards are drawn randomly in each draw out of the remaining cards. So, it is the situation of a hypergeometric distribution. In usual notations, we are given

G = 12 (4 jack + 4 queen + 4 kings), B = 52 – G = 40, n = 39. Let X denotes the number of jack or queen or king out of these 39 selected cards, then the expected value of X is given by

$$E(X) = \frac{nG}{(G+B)} = \frac{39 \times 12}{12+40} = \frac{39 \times 12}{52} = 9.$$

Example 3: **Application in the field of Medicine**: If the probability that an individual suffers a bad reaction from an injection of a given serum is 0.001, using an appropriate probability distribution find the probability that out of 5000 individuals

(i) exactly 10,

(ii) more than 10

individuals suffer from bad reaction. Assume that each individual has almost equal chance of a bad reaction.

Solution: Here probability of success is small and sample size is large. So, under the assumption that each individual has almost equal chance of bad reaction appropriate probability distribution is Poisson distribution. So, in usual notations, we are given

n = 5000, p = 0.001. So, if X counts the "Number of individuals suffering from bad reaction", then $X \sim \text{Pois}(\lambda)$, where

$\lambda = np = 5000(0.001) = 5.$

We know that PMF of Poisson distribution is given by

$$p_X(x) = \mathcal{P}(X = k) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{\lfloor k}, & k = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

(i) Required probability =

$$\mathcal{P}[X = 10] = \frac{e^{-5}5^{10}}{\lfloor 10} = 0.01813279 \begin{bmatrix}\text{Using scientific} \\ \text{calculator}\end{bmatrix}$$

You can also verify this answer using R as a calculator as follows.

```
> (5^10)*exp(-5)/(factorial(10))
[1] 0.01813279
```

You can also obtain it using dpois() function as mentioned in (11.14). Screenshot of R code with output is shown as follows.

```
> dpois(x = 10, lambda = 5)
[1] 0.01813279
```

If you do not specify the names 'x' and 'lambda' of the arguments of dpois() function then R matches them by their positions. For example, previous output can also be obtained as follows.

```
> dpois(10,5)
[1] 0.01813279
```

Remember in Term End Exam (TEE) of this course, you have to obtain it using scientific calculator which is allowed in your exam while in Lab exam of the course MSTL-015, you have to obtain it using R programming language.

(ii) Required probability =

$$\mathcal{P}[X > 10] = 1 - \mathcal{P}[X \le 10] = 1 - \left[\sum_{x=0}^{10} \mathcal{P}[X = x]\right]$$

$$= 1 - \left[\sum_{x=0}^{10} \frac{e^{-5}5^x}{\lfloor x}\right] = 1 - 0.9863047 \begin{bmatrix}\text{Using scientific} \\ \text{calculator}\end{bmatrix}$$

$$= 0.0136953$$

You can also obtain it using ppois() function as mentioned in (11.14). Screenshot of R code with output is shown as follows.

```
> ppois(q = 10, lambda = 5)
[1] 0.9863047
```

If you want final answer then, you can run following R code.

```
> 1 - ppois(q = 10, lambda = 5)
[1] 0.01369527
```

Example 4: In a city 500 births of babies take place each month out of which 1 in 60 birth is of a twin. Find the probability that in the next month there will be 8 twins

Solution: Here probability of a twin is 1/60 which is small and number of births is 500 which is large. So, Poisson distribution is an appropriate probability distribution to obtain required probability. In usual notations, we given

n = 500, p = 1/60 and so $\lambda = np = 500(1/60) = 25/3.$ Let the random variable X denote the number of twin births in the next month then $X \sim Pois(\lambda = 25/3).$ So, required probability is given by

$$\mathcal{P}(X=8) = \frac{e^{-25/3}(25/3)^8}{\underline{|8}} = 0.1386465$$

Example 5: The Powerball game is a game where, you have to choose 5 white integers among 1 to 55 and 1 red integer among 1 to 42. In this game there are various prizes like, you choose (i) all the five wining white balls and Powerball, where wining red ball in known as Powerball (ii) only five wining white balls (iii) only four wining white balls, and so on. Find the probability of only selecting 3 wining white balls.

Solution: Let X denote the number of choosing wining white balls. Since all the five balls among the 55 white balls are selected simultaneously. So, we will use hypergeometric distribution to obtain required probability. In usual notations, we are given

G = 5, B = 50, n = 5. So, $\mathcal{P}(X=3) = \frac{\binom{5}{3}\binom{50}{2}}{\binom{55}{5}} = 0.003521369.$

But as per the rules of the game to play this game one also has to choose a red integer. But we are interested in the probability of both getting only 3 winning white integers. It means, we have to select non-winning red integer and it is given by 41/42. So, using multiplication rule for independent events final probability is given by

$$0.003521369 \times \frac{41}{42} = 0.003437527.$$

Now, you can try the following two Self-Assessment Questions.

SAQ 1

A lot of 50 units contains 4 defective units. An engineer inspects 3 randomly selected units from the lot. He/She accepts the lot if all the three units are found in good condition, otherwise all the remaining units are inspected. Find the probability that the lot is accepted without further inspection.

SAQ 2

If X ~ Pois(2) and Y ~ Pois(3) are two independent random variables, then find $\mathcal{P}(X+Y<4)$.

11.7 SUMMARY

A brief summary of what, we have covered in this unit is given as follows:

- **Definition of Discrete Uniform Distribution**: If X ~ B(n, p) such that n approaches to infinity, p is small and $np=\lambda$ is a fixed finite number then binomial PMF converges to the PMF given by

$$p_X(x)=\mathcal{P}(X=k)=\begin{cases}\dfrac{e^{-\lambda}\lambda^k}{\underline{|k}}, & k=0,1,2,3,\ldots\\ 0, & \text{otherwise}\end{cases}$$

  This is the PMF of Poisson distribution with parameter λ and is denoted by writing $X\sim \text{Pois}(\lambda)$.

- **CDF of Discrete Uniform Distribution**: Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x)=\mathcal{P}(X\le x)=\begin{cases}0, & \text{if } x<0\\ \displaystyle\sum_{k=0}^{[x]}\frac{e^{-\lambda}\lambda^k}{\underline{|k}}, & \text{if } 0\le x<\infty\\ 1, & \text{if } x\to\infty\end{cases}$$

- **Summary measures of Poisson distribution**

Name of measure	Formula	Name of measure	Formula
Mean	λ	MGF	$SD(X)=\sigma=\sqrt{\dfrac{(b-a+1)^2-1}{12}}$
Variance	λ	Skewness	$3+\dfrac{1}{\lambda}$
Standard deviation	$\sqrt{\lambda}$	Kurtosis	$3+\dfrac{1}{\lambda}$

- **Definition of Hypergeometric Distribution**: Suppose in a bag there are total N items out of which G items are good and remaining B = N – G items are bad. We draw n items randomly from the given bag having N items one by one without replacement. If we are interested in the probability of getting g good items out of n = g + b = drawn items where b represents number of bad items then it is given by $\dfrac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}$.

  So, PMF, $p_X(x)$ of hypergeometric distribution is given by

$$p_X(x) = \mathcal{P}\left(X = x\right) = \begin{cases} \dfrac{\binom{G}{x}\binom{B}{n-x}}{\binom{N}{n}}, & \text{if } x = 0, 1, 2, \ldots, \min\{G, n\} \\ 0, & \text{otherwise} \end{cases}$$

If random variable X has this PMF, then we say that it follows hypergeometric distribution with parameters G, B and n and is denoted by writing $X \sim \text{HGeom}(G, B, n)$.

- **CDF of Bernoulli Distribution**: Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}\left(X \le x\right) = \begin{cases} 0, & \text{if } x < 0 \\ \dfrac{\sum_{x=0}^{[x]}\binom{G}{x}\binom{B}{n-x}}{\binom{N}{n}}, & \text{if } 0 \le x < \min\{G, n\} \\ 1, & \text{if } x \ge \min\{G, n\} \end{cases}$$

- **Summary measures of Hypergeometric distribution**

Name of measure	Formula	Name of measure	Formula
Mean	$\dfrac{nG}{(G+B)}$	Standard deviation	$\sqrt{\dfrac{nGB(G+B-n)}{(G+B)^2(G+B-1)}}$
Variance	$\dfrac{nGB(G+B-n)}{(G+B)^2(G+B-1)}$		

11.8 TERMINAL QUESTIONS

1. Let us suppose that in a lake there are 1200 fishes. 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, 170 will be having red spots.

2. For a Poisson distribution, it is given that $\mathcal{P}\left(X=1\right) = \mathcal{P}\left(X=2\right)$, find the value of mean and variance of distribution. Hence find $\mathcal{P}\left(X=0\right)$ and $\mathcal{P}\left(X=3\right)$.

11.9 SOLUTIONS/ANSWERS

Self-Assessment Questions (SAQs)

1. It is a situation of hypergeometric distribution where in usual notations, we are given G = 46, B = 4 and n = 3. So, required probability is given by

$$\mathcal{P}\begin{bmatrix}\text{none of the 3 randomly selected} \\ \text{units is found defective}\end{bmatrix} = \frac{\binom{4}{0}\binom{46}{3}}{\binom{50}{3}} = 0.7744898.$$

2. To solve this problem, you have to remember one result of Poisson distribution, if $X \sim \text{Pois}(\lambda_1)$ and $Y \sim \text{Pois}(\lambda_2)$, are two independent random Poisson variables, then then $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$. In this problem, we have $\lambda_1 = 2$ and $\lambda_2 = 3$, so $\lambda_1 + \lambda_2 = 2 + 3 = 5$. Hence, Z = X + Y ~ Pois(5). Using PMF of Poisson distribution required probability is given by

$$\mathcal{P}(Z < 4) = \mathcal{P}(Z = 0) + \mathcal{P}(Z = 1) + \mathcal{P}(Z = 2) + \mathcal{P}(Z = 3)$$

$$= \frac{e^{-5}(5)^0}{\lfloor 0} + \frac{e^{-5}(5)^1}{\lfloor 1} + \frac{e^{-5}(5)^2}{\lfloor 2} + \frac{e^{-5}(5)^3}{\lfloor 3} = 0.2650259$$

Terminal Questions

1. In the lake there are only two types of fish with red spot and without red spot. Also, fish are caught simultaneously. So, it is a situation of the hypergeometric distribution. If we call red spot fish as good/success then in usual notations, we are given

   G = 500, B = 1200 – 500 = 700, n = 400. So, required probability is given by

$$\mathcal{P}(X = 170) = \frac{\binom{500}{170}\binom{700}{230}}{\binom{1200}{400}} = 0.0453958$$

2. Let $X \sim \text{Pois}(\lambda)$ so using PMF of Poisson distribution, we have

$$\mathcal{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{\lfloor x}; \; x = 0, 1, 2, 3, \ldots$$

$$\therefore \; \therefore \; \mathcal{P}(X = 1) = \mathcal{P}(X = 2) \Rightarrow \frac{e^{-\lambda}\lambda^1}{\lfloor 1} = \frac{e^{-\lambda}\lambda^2}{\lfloor 2} \Rightarrow 2\lambda = \lambda^2 \Rightarrow \lambda^2 - 2\lambda = 0$$

$$\Rightarrow \lambda(\lambda - 2) = 0 \Rightarrow \lambda = 0, 2.$$

Rejecting λ = 0 because in that case random variable will become a constant random variable. So, let λ = 2.

We know that both mean and variance of Poisson distribution are λ. Hence, mean = variance = 2.

Now, $\mathcal{P}(X = 0) = \frac{e^{-\lambda}\lambda^0}{\lfloor 0} = e^{-2} = 0.1353353$, and

$$\mathcal{P}(X = 3) = \frac{e^{-\lambda}\lambda^3}{\lfloor 3} = \frac{e^{-2}(2)^3}{6} = 0.180447. \quad \text{[Using scientific calculator]}$$

UNIT 12

GEOMETRIC AND NEGATIVE BINOMIAL DISTRIBUTIONS

Structure

12.1 INTRODUCTION

We know that support of a discrete random variable either may be a finite set or may be a countably infinite set of values. In Units 9 and 10 we studied four probability distributions namely (i) discrete uniform (ii) Bernoulli (iii) Binomial and (iv) Multinomial distributions each of which has a finite support. In Unit 11, we studied Poisson distribution which has countably infinite support. Another two discrete probability distributions which have countably infinite support are geometric and negative binomial distributions which will be discussed in this unit. In Sec. 12.2, we will discuss PMF and CDF of geometric distribution while in Sec. 12.3, we will discuss MGF and some other summary measures of this probability distribution.

In Secs. 12.4 and 12.5, we will do similar studies about negative binomial distribution. Some applications of these distributions are discussed in Sec. 12.6.

What we have discussed in this unit is summarised in Sec. 12.7. Self-Assessment Questions (SAQs) have been given in some sections which are generally based on the content discussed in that section. But to give you a good practice of what we have discussed in this unit some more questions based on the entire unit are given in Sec. 12.8 under the heading Terminal Questions. Due to the reason mentioned in Sec. 1.1 of Unit 1 of this course, solutions of all the SAQs and Terminal Questions are given in Sec. 12.9.

In the next unit, you will do a similar study about two continuous probability distributions known as continuous uniform and exponential distributions.

Expected Learning Outcomes

After completing this unit, you should be able to:

- ❖ identify situations where you can apply geometric and negative binomial distributions;
- ❖ define PMF, CDF, MGF and some summary measures of geometric and negative binomial distributions; and
- ❖ apply geometric and negative binomial distributions to solve problems based on these two probability distributions.

12.2 STORY, DEFINITION, PMF AND CDF OF GEOMETRIC DISTRIBUTION

In Sec. 10.2 of the Unit 10, you have studied binomial distribution. Recall that in a binomial distribution, we have

- The number of trials 'n' is finite in numbers. … (12.1)
- All the 'n' trials are independent. … (12.2)
- Probability of success p remains constant in each trial. That is, p does not change from trial to trial. … (12.3)

Recall that in binomial distribution, we perform the experiment n times (**a fixed number**) and are interested in the probability of getting x (**variable**) successes out of n trials where x = 0, 1, 2, 3, …, n. But if we perform the experiment till we get the first success then, we can define two random variables as follows.

First Random Variable: Let X be the random variable which counts the number of failures before the first success.

Second Random Variable: Let Y be the random variable which counts the number of trials till the first success including the first success.

The probability distributions of both the random variables X and Y are known as geometric distribution. We call it geometric because successive probabilities form geometric progression refer to (12.7). We will proceed with the first random variable.

We are following the same notations as we used in the binomial distribution. So, probability of success and failure will be denoted by p and q = 1 – p. Suppose we get the first success in x^{th} trial then all the first x – 1 trials will be failure. Since trials are independent refer to (12.2) so, probability of getting the first success in x^{th} trial is given by

$$\mathcal{P}(X = x) = \underbrace{qqq\ldots q}_{x-\text{ times}}p = q^x p = (1-p)^x p \qquad \ldots (12.4)$$

But the first success may occur in the first trial or second trial or third trial, and so on. If we get the first success in the first trial then the number of failures before the first success will be 0. If we get the first success in the second trial

then the number of failure before the first success will be 1. If we get the first success in the third trial then the number of failures before the first success will be 2, and so on. So, values of x in (12.4) may be 0 or 1 or 2 or 3 ….
Hence, finally (12.4) can be written as

$$\mathcal{P}(X = x) = \underbrace{qqq\ldots q}_{x-\text{times}}p = q^x p = (1-p)^x p, \ \ x = 0, 1, 2, 3, 4, \ldots \qquad \ldots (12.5)$$

Now, we can discuss the story of the geometric distribution.

Story of Geometric Distribution: If we perform Bernoulli trials till, we get the first success then both the random variables which count the number of:

(i) failures before the first success; and

(ii) trials till the first success including the first success

forms a probability distribution known as a geometric distribution.

Definition and PMF of Geometric Distribution: If we perform Bernoulli trials till, we get the first success and the random variable X counts the number of failures before the first success then the PMF of X is given by

$$p_X(x) = \mathcal{P}(X = x) = \begin{cases} (1-p)^x p, & x = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (12.6)$$

Note that successive probabilities of the PMF given by (12.6) are $p, \ (1-p)p, \ (1-p)^2 p, \ (1-p)^3 p, \ldots$ which form a geometric progression (GP) with the first term p and the common ratio 1 – p because of this reason PMF given by (12.6) is known as PMF of a geometric distribution with single parameter p and is denoted by writing X ~ Geom(p). … (12.7)

Like other distributions, we read $X \sim \text{Geom}(p)$ as X is distributed as a geometric distribution with parameter p. Or we read it as X follows a geometric distribution with parameter p. … (12.8)

Since the statistical software used for hands-on practice for the lab courses of this programme is R programming language, so, you should also know its notation in R. In R notation that is used for geometric distribution is geom(prob) in the stats package, where prob represents the value of the parameter p of the geometric distribution. In fact, there are four functions for geometric distribution namely dgeom(x, prob, …), pgeom(q, prob, …), qgeom(p, prob, …), and rgeom(n, prob, …). We have already explained the meaning of these functions in Unit 9. … (12.9)

Let us check the **validity of the PMF of the Geometric distribution**.

We have to check two conditions non-negativity and normality refer to (4.62).

(1) **Non-negativity**: Since $0 < p \le 1 \Rightarrow (1-p)^x p \ge 0 \ \ \forall \ x = 0, 1, 2, 3, \ldots$ hence

$$p_X(x) = \mathcal{P}(X = x) \ge 0 \qquad \ldots (12.10)$$

(2) **Normality**: Let us obtain the sum of all probabilities of geometric distribution.

$$\sum_{x=0}^{\infty}(1-p)^x p = \sum_{x=0}^{\infty} q^x p = p\sum_{x=0}^{\infty} q^x = p\left(q^0 + q^1 + q^2 + q^3 + \ldots\right)$$

$$= p\left(1 + q + q^2 + q^3 + \ldots\right)$$

$$= p\frac{1}{1-q}\left[\begin{array}{l}\because \text{ Sum of infinite GP } a + ar + ar^2 + ar^3 + \ldots \\ = \dfrac{a}{1-r},\ \text{where } |r| < 1\end{array}\right]$$

$$= \frac{p}{p} = 1\ \left[\because 1-q = p\right] \qquad \ldots (12.11)$$

This proves that sum of all probabilities of geometric distribution is 1. So, we can say that PMF of the geometric random variable is a valid PMF.

Now, we define the CDF of a geometric distribution.

CDF of Geometric Distribution: If X ~ Geom(p) then PMF of X is given by

$$p_X(x) = \mathcal{P}\left(X = x\right) = \begin{cases}(1-p)^x p, & x = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise}\end{cases}$$

Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}\left(X \le x\right) = \begin{cases}0, & \text{if } x < 0 \\ \displaystyle\sum_{k=0}^{[x]}(1-p)^x p = p\frac{1-(1-p)^{[x]+1}}{1-(1-p)} = 1 - q^{[x]+1}, & \text{if } x \ge 0\end{cases}$$

$$\left[\begin{array}{l}\because \text{ Sum of } n \text{ terms of a GP } a + ar + ar^2 + ar^3 + \ldots = \dfrac{a(1-r^n)}{1-r},\ \text{where } |r| < 1 \\ \text{Here, } a = p,\ r = 1-p \text{ and } n = [x]+1\end{array}\right]$$

or

$$F_X(x) = \mathcal{P}\left(X \le x\right) = \begin{cases}0, & \text{if } x < 0 \\ 1-(1-p)^{[x]+1}, & \text{if } x \ge 0\end{cases} \qquad \ldots (12.12)$$

Let us do one example.

Example 1: A biased coin which has the probability of getting a head as 1/4 and a tail as 3/4 is tossed till, we get the first head. If X counts the number of failures before the first head, then obtain the PMF and CDF of X and also plot the PMF and CDF of X.

Solution: If we define getting a head as a success and getting a tail as a failure then the random variable X which counts the number of failures before the first head follows a geometric distribution with parameter p = 1/4. So, the PMF of X is given by

$$p_X(x) = \mathcal{P}\left(X = x\right) = \begin{cases}\left(\dfrac{3}{4}\right)^x\left(\dfrac{1}{4}\right), & x = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise}\end{cases} \qquad \ldots (12.13)$$

and the CDF of X is given by

$$F_X(x) = \mathcal{P}\left(X \le x\right) = \begin{cases}0, & \text{if } x < 0 \\ 1-\left(\dfrac{3}{4}\right)^{[x]+1}, & \text{if } x \ge 0\end{cases} \qquad \ldots (12.14)$$

Probabilities for X = 0, 1, 2, 3, …, 10 are given by

$$\mathcal{P}(X=0)=\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)=0.25 \qquad \mathcal{P}(X=6)=\left(\frac{3}{4}\right)^6\left(\frac{1}{4}\right)=0.04449463$$

$$\mathcal{P}(X=1)=\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)=0.1875 \qquad \mathcal{P}(X=7)=\left(\frac{3}{4}\right)^7\left(\frac{1}{4}\right)=0.03337097$$

$$\mathcal{P}(X=2)=\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)=0.140625 \qquad \mathcal{P}(X=8)=\left(\frac{3}{4}\right)^8\left(\frac{1}{4}\right)=0.02502823$$

$$\mathcal{P}(X=3)=\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)=0.1054688 \qquad \mathcal{P}(X=9)=\left(\frac{3}{4}\right)^9\left(\frac{1}{4}\right)=0.01877117$$

$$\mathcal{P}(X=4)=\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)=0.07910156 \qquad \mathcal{P}(X=10)=\left(\frac{3}{4}\right)^{10}\left(\frac{1}{4}\right)=0.01407838$$

$$\mathcal{P}(X=5)=\left(\frac{3}{4}\right)^5\left(\frac{1}{4}\right)=0.05932617 \qquad \mathcal{P}(X=11)=\left(\frac{3}{4}\right)^{11}\left(\frac{1}{4}\right)=0.01055878$$

Now, using (12.14) CDF of X is given by

$$F_X(x)=\mathcal{P}(X\le x)=\begin{cases}0, & \text{if } x<0\\ 0.25, & \text{if } 0\le x<1\\ 0.4375, & \text{if } 1\le x<2\\ 0.578125, & \text{if } 2\le x<3\\ 0.6835938, & \text{if } 3\le x<4\\ 0.7626953, & \text{if } 4\le x<5\\ 0.8220215, & \text{if } 5\le x<6\\ 0.8665161, & \text{if } 6\le x<7\\ 0.8998871, & \text{if } 7\le x<8\\ 0.9249153, & \text{if } 8\le x<9\\ 0.9436865, & \text{if } 9\le x<10\\ 0.9577649, & \text{if } 10\le x<11\\ \vdots & \end{cases} \qquad \dots (12.15)$$

PMF and CDF are plotted in Fig. 12.1 (a) and (b) respectively as follows.

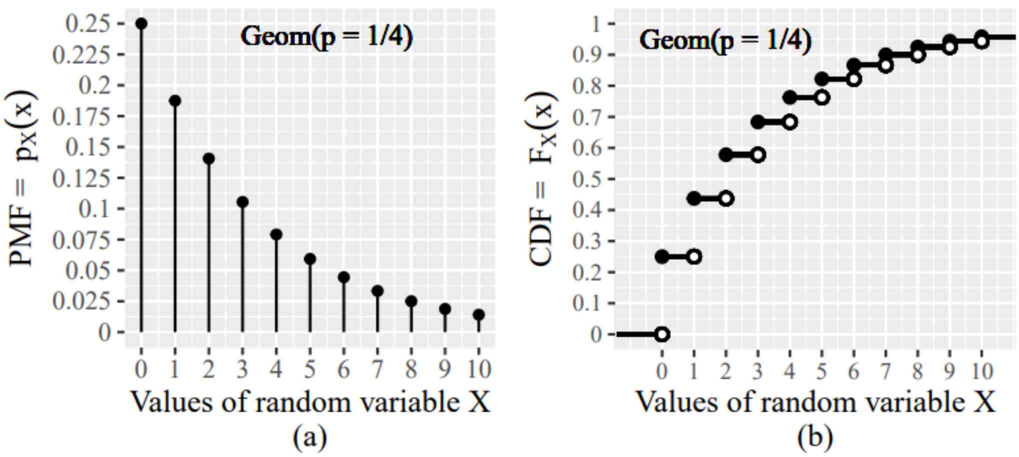


Fig. 12.1: Visualisation of (a) PMF (b) CDF of the Geom(1/4) discussed in Example 1

12.3 MGF AND OTHER SUMMARY MEASURES OF GEOMETRIC DISTRIBUTION

In the previous section, you have studied PMF and CDF of geometric distribution. In this section, we want to obtain MGF and some other summary measure of geometric distribution like mean and variance. Let us first obtain MGF of geometric distribution. We will obtain the MGF of geometric distribution using the definition of MGF, you may refer to (7.48).

Calculation of MGF

$$M_X(t) = E\left(e^{tX}\right),\ t \in \mathbb{R}$$

$$= \sum_{x=0}^{\infty} e^{tx} q^x p = p\sum_{x=0}^{\infty}\left(qe^t\right)^x = p\left[\left(qe^t\right)^0 + \left(qe^t\right)^1 + \left(qe^t\right)^2 + \left(qe^t\right)^3 + \ldots\right]$$

$$= p\left[1 + qe^t + \left(qe^t\right)^2 + \left(qe^t\right)^3 + \ldots\right]$$

$$= p\left(\frac{1}{1-qe^t}\right) \quad \left[\begin{array}{l}\because \text{ sum of infinite GP } a + ar + ar^2 + \ldots = \dfrac{a}{1-r} \\ \text{Provided } \left|qe^t\right| < 1 \Rightarrow \left|(1-p)e^t\right| < 1 \Rightarrow t < \ln\left(\dfrac{1}{1-p}\right)\end{array}\right]$$

$$= \frac{p}{1-(1-p)e^t},\quad t < \ln\left(\frac{1}{1-p}\right)$$

$$M_X(t) = \frac{p}{1-(1-p)e^t},\ t < \ln\left(\frac{1}{1-p}\right) \quad \ldots (12.16)$$

Calculation for Mean

$$\text{Expected value} = E(X) = \sum_{x=0}^{\infty} x\, q^x\, p = p\sum_{x=1}^{\infty} x\, q^x \left[\begin{array}{l}\because \text{ when } x = 0, \text{ then we get} \\ \text{value of } x\,q^x\,p = 0\end{array}\right]$$

$$= p\sum_{x=1}^{\infty} x\, q\, q^{x-1} = pq\sum_{x=1}^{\infty} x\, q^{x-1} = pq\sum_{x=1}^{\infty}\frac{d}{dq}\left(q^x\right) \quad \left[\because \frac{d}{dx}\left(q^x\right) = x\,q^{x-1}\right]$$

$$= pq\frac{d}{dq}\left(\sum_{x=1}^{\infty} q^x\right) \quad [\because \text{ summation respect differentiation}]$$

$$= pq\frac{d}{dq}\left(q + q^2 + q^3 + q^4 + \ldots\right)$$

$$= pq\frac{d}{dq}\left(\frac{q}{1-q}\right) \quad \left[\begin{array}{l}\because \text{ sum of infinite GP,} \\ a + ar + ar^2 + ar^3 + \ldots = \dfrac{a\left(1-r^n\right)}{1-r}\end{array}\right]$$

$$= pq\left[\frac{(1-q)(1) - q(-1)}{(1-q)^2}\right] = pq\left[\frac{1}{p^2}\right] \quad [\because 1 - q = p]$$

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

Hence, mean of X = expected value of $X = E(X) = \dfrac{1-p}{p}$. … (12.17)

Calculation for Variance and Standard Deviation

$$E\left(X(X-1)\right)=\sum_{x=0}^{\infty}x(x-1)q^x p=p\sum_{x=2}^{\infty}x(x-1)q^2q^{x-2}\left[\begin{array}{l}\because \text{ when } x=0,1 \text{ then}\\ \text{value of } x(x-1)q^x p=0\end{array}\right]$$

$$=pq^2\sum_{x=2}^{\infty}x(x-1)q^{x-2}=pq^2\sum_{x=2}^{\infty}\frac{d^2}{dq^2}\left(q^x\right)=pq^2\frac{d^2}{dq^2}\left(\sum_{x=2}^{\infty}q^x\right)$$

$$=pq^2\frac{d^2}{dq^2}\left(q^2+q^3+q^4+q^5+\ldots\right)=pq^2\frac{d^2}{dq^2}\left(\frac{q^2}{1-q}\right)$$

$$=pq^2\frac{d}{dq}\left(\frac{(1-q)(2q)-q^2(-1)}{(1-q)^2}\right)$$

$$=pq^2\frac{d}{dq}\left(\frac{2q-q^2}{(1-q)^2}\right)=pq^2\left(\frac{(1-q)^2(2-2q)-\left(2q-q^2\right)2(1-q)(-1)}{(1-q)^4}\right)$$

$$=pq^2\left(\frac{(1-q)(2-2q)+2\left(2q-q^2\right)}{(1-q)^3}\right)$$

$$=pq^2\left(\frac{2-2q-2q+2q^2+4q-2q^2}{p^3}\right)=q^2\left(\frac{2}{p^2}\right)=\frac{2q^2}{p^2}$$

$$E\left(X(X-1)\right)=\frac{2q^2}{p^2} \qquad \ldots (12.18)$$

$$E\left(X^2\right)=E\left(X(X-1)+X\right)=E\left(X(X-1)\right)+E\left(X\right) \quad \left[\begin{array}{l}\text{Using addition theorem of}\\ \text{expectation refer to (7.29)}\end{array}\right]$$

$$=\frac{2q^2}{p^2}+\frac{q}{p} \quad \left[\text{Using (12.18) and (12.17)}\right]$$

$$\Rightarrow E\left(X^2\right)=\frac{2q^2}{p^2}+\frac{q}{p} \qquad \ldots (12.19)$$

Using (7.63) variance of any random variable X is given by

$$\text{Variance of the geometric distribution}=\mu_2=E\left(X^2\right)-\left(E(X)\right)^2$$

$$=\frac{2q^2}{p^2}+\frac{q}{p}-\left(\frac{q}{p}\right)^2 \left[\begin{array}{l}\text{Using (12.19)}\\ \text{and (12.17)}\end{array}\right]$$

$$=\frac{q^2}{p^2}+\frac{q}{p}=\frac{q^2+pq}{p^2}=\frac{q(q+p)}{p^2}$$

$$=\frac{q}{p^2}=\frac{1-p}{p^2} \qquad \left[\because p+q=1\right]$$

Hence, variance of the geometric distribution $=\dfrac{1-p}{p^2}$. … (12.20)

We know that standard deviation of X is positive square root of variance of X.

Hence, $SD(X)=\sqrt{\text{Variance of } X}=\dfrac{\sqrt{1-p}}{p}$. … (12.21)

Similarly, we can obtain other measures of our interest. But keeping applied nature of the programme in view, we are not focusing on proof of each measure. Some commonly used summary measures of geometric distribution are shown in Table 12.1 given as follows.

Table 12.1: Summary measures of geometric distribution

Name of measure	Formula	Name of measure	Formula
Mean	$\frac{1-p}{p}$	MGF	$\frac{p}{1-(1-p)e^t}$
Variance	$\frac{1-p}{p^2}$	Skewness	$\frac{2-p}{\sqrt{1-p}}$
Standard deviation	$\frac{\sqrt{1-p}}{p}$	Kurtosis	$6+\frac{p^2}{1-p}$

12.4 STORY, DEFINITION, PMF AND CDF OF NEGATIVE BINOMIAL DISTRIBUTION

In geometric distribution, we perform the experiment till, we get the first success but instead of the first success if we perform the experiment till r^{th} success then distribution of the random variable which counts the number of failures before the r^{th} success is known as negative binomial distribution. So, we can say that negative binomial distribution is the generalisation of geometric distribution or geometric distribution is the particular case of negative binomial distribution when r = 1. … (12.22)

Now, we can write the story of negative binomial distribution as follows.

Story of Negative Binomial Distribution: If we perform Bernoulli trial where probability of success is p and probability of failure is q = 1 – p till, we get r^{th} success then distribution of the random variable X which counts the number of failures before the r^{th} success is known as negative binomial distribution. So, negative binomial distribution models the distribution of Bernoulli trials where we are interested in the number of failures before the r^{th} success. … (12.23)

Let us obtain the probability of getting r^{th} success in $(x + r)^{th}$ trial where x represents the number of failures before the r^{th} success. This can happen only when $(x+r)^{th}$ trial is a success, but the remaining (r – 1) successes occur in the $x+r-1$ trials before the $(x + r)^{th}$ trial. But each trial is a Bernoulli trial with the same probability of success p, so the probability of getting (r – 1) successes out of (x + r – 1) trials can be obtained using binomial distribution Bin(x + r – 1, p) and is given by

$$\mathcal{P}\begin{pmatrix}\text{getting } r-1 \text{ successes}\\ \text{out of } x+r-1 \text{ trials}\end{pmatrix}=\binom{x+r-1}{r-1}p^{r-1}(1-p)^{x+r-1-(r-1)}$$

$$=\binom{x+r-1}{r-1}p^{r-1}(1-p)^{x}$$

Now, required probability is given by

$$\mathcal{P}\begin{pmatrix}\text{getting x failures}\\ \text{before } r^{th} \text{ success}\end{pmatrix} = \mathcal{P}\left(\begin{Bmatrix}\text{getting x} - 1 \text{ failures in}\\ \text{the first x} + \text{r} - 1 \text{ trials}\end{Bmatrix} \cap \begin{Bmatrix}\text{getting } r^{th} \text{ success}\\ \text{in } (x+r)^{th} \text{ trial}\end{Bmatrix}\right)$$

$$= \mathcal{P}\begin{pmatrix}\text{getting x} - 1 \text{ failures in}\\ \text{the first x} + \text{r} - 1 \text{ trials}\end{pmatrix} \mathcal{P}\begin{pmatrix}\text{getting } r^{th} \text{ success}\\ \text{in } (x+r)^{th} \text{ trial}\end{pmatrix}$$

$$[\because \text{both the events are independent}]$$

$$= \binom{x+r-1}{r-1} p^{r-1} (1-p)^x p$$

$$= \binom{x+r-1}{r-1} p^r (1-p)^x \qquad \ldots (12.24)$$

Now, we define negative binomial distribution as follows.

Definition and PMF of Negative Binomial Distribution: If we perform Bernoulli trial where probability of success is p and probability of failure is q = 1 – p till, we get r^{th} success then PMF of the random variable X which counts the number of failures before the r^{th} success is given by

$$p_X(x) = \mathcal{P}(X = x) = \begin{cases} \binom{x+r-1}{r-1} p^r (1-p)^x, & \text{if } x = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (12.25)$$

If random variable X has PMF given by (12.25), then, we say that it follows negative binomial distribution with parameters r and p and is denoted by writing $X \sim \text{NBin}(r, p)$. … (12.26)

We read it as X follows negative binomial distribution with parameters r and p. … (12.27)

Since the statistical software used for hands-on practice for the lab courses of this programme is R programming language, so, you should also know its notation in R. In R notation that is used for negative binomial distribution is nbinom(size, prob, …) in the stats package, where size represents the number of success and p represents the probability of success. In fact, there are four functions for negative binomial distribution namely dnbinom(x, size, prob, …), pnbinom(q, size, prob, …), qnbinom(p, size, prob, …) and rnbinom(n, size, prob, …). We have already explained the meaning of these functions in Unit 9. ... (12.28)

Let us check the **validity of the PMF of the negative binomial distribution**.

In checking the validity of PMF of negative distribution, we will use Vandermonde's Identity. So, let first state and prove it.

Vandermonde's Identity: For positive integer n and non-negative integer x such that $x \le n$ prove that $\binom{n}{x} = \binom{n}{n-x}$. … (12.29)

Proof: Suppose a bag has n items. Each time, we select x items out of n items then automatically in the bag, we are left with n – x items. So, the number of ways of selecting x items out of n is equal to selecting n – x items out of n.

Thus, we have $\binom{n}{x} = \binom{n}{n-x}, \forall\ x, 0 \le x \le n.$

Now, we can check the **validity of the PMF of the negative binomial distribution**.

We have to check two conditions non-negativity and normality refer to (4.62).

(1) Non-negativity: Since $0 < p \le 1$ and $\binom{x+r-1}{r-1} > 0$, so

$$\binom{x+r-1}{r-1} p^r (1-p)^x \ge 0, \;\; \forall \; x = 0, 1, 2, 3, \ldots$$

$$\Rightarrow p_X(x) = \mathcal{P}\left(X = x\right) \ge 0 \;\; \forall \; x = 0, 1, 2, 3, \ldots \qquad \ldots (12.30)$$

(2) **Normality**: Required to prove

$$\sum_{x=0}^{\infty} p_X(x) = \sum_{x=0}^{\infty} \mathcal{P}\left(X = x\right) = \sum_{x=0}^{\infty} \binom{x+r-1}{r-1} p^r (1-p)^x = 1 \qquad \ldots (12.31)$$

Using (12.29), we have

$$\binom{x+r-1}{r-1} = \binom{x+r-1}{x+r-1-(r-1)}$$

$$\Rightarrow \binom{x+r-1}{r-1} = \binom{x+r-1}{x}$$

$$= \frac{\underline{|x+r-1}}{\underline{|x}\,\underline{|x+r-1-x}} \quad \left[\because \binom{n}{x} = \frac{\underline{|n}}{\underline{|x}\,\underline{|n-x}}\right]$$

$$= \frac{\underline{|r+x-1}}{\underline{|x}\,\underline{|r-1}}$$

$$= \frac{\left(r+(x-1)\right)\left(r+(x-2)\right)\ldots(r+1)(r)\underline{|r-1}}{\underline{|x}\,\underline{|r-1}}$$

$$= \frac{(r+x-1)(r+x-2)\ldots(r+1)r}{\underline{|x}}$$

$$= \frac{r(r+1)(r+2)\ldots(r+x-2)(r+x-1)}{\underline{|x}} \quad \begin{bmatrix}\text{Writing terms in}\\ \text{reverse order}\end{bmatrix}$$

Taking – 1 common from each of the x terms in the numerator, we have

$$\binom{x+r-1}{r-1} = \frac{(-1)^x(-r)(-r-1)(-r-2)\ldots\{-r-(x-2)\}\{-r-(x-1)\}}{\underline{|x}}$$

We know that the symbol $\binom{n}{x}$ stands for ${}^nC_x$ which represents the number of combinations of n things taken x at a time if n and x are positive integers and is equal to $\frac{n(n-1)(n-2)\ldots\{n-(x-1)\}}{\underline{|x}}$. We may also use the symbol $\binom{n}{x}$ if n is any real number but, in that case, though it will not have the interpretation as mentioned in the case of positive integers. Yet it is equal to $\frac{n(n-1)(n-2)\ldots\{n-(n-x)\}}{\underline{|x}}$. So, we have

$$\binom{x+r-1}{r-1} = (-1)^x \binom{-r}{x} \qquad \ldots (12.32)$$

Using (12.32) in (12.31), we have

$$\sum_{x=0}^{\infty} p_X(x) = \sum_{x=0}^{\infty} \mathcal{P}(X=x) = \sum_{x=0}^{\infty} (-1)^x \binom{-r}{x} p^r (1-p)^x$$

$$= p^r \sum_{x=0}^{\infty} \binom{-r}{x} (-1+p)^x \qquad \left[\because (-1)^x (1-p)^x = (-1+p)^x\right]$$

$$= p^r \sum_{x=0}^{\infty} \binom{-r}{x} (-1+p)^x (1)^{-r-x} \qquad \left[\because (1)^{-r-x} = 1\right]$$

$$= p^r (-1+p+1)^{-r} \qquad \left[\begin{array}{l}\because \sum_{x=0}^{\infty} \binom{-m}{x} a^x (1)^{-m-x} = (a+1)^{-m} \\ \text{where m is a positive integer}\end{array}\right]$$

$$= p^r (p)^{-r} = p^{r-r} = p^0 = 1 \qquad \ldots (12.33)$$

This proves (12.31). Hence, sum of all probabilities of negative binomial distribution is 1.

Hence, PMF defined by (12.25) of negative binomial distribution is a valid PMF.

Now, we define CDF of negative binomial distribution.

CDF of Negative Binomial Distribution: If $X \sim \text{NBin}(r, p)$, then PMF of X is given by

$$p_X(x) = \mathcal{P}(X=x) = \begin{cases} \binom{x+r-1}{r-1} p^r (1-p)^x, & \text{if } x = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}(X \le x) = \begin{cases} 0, & \text{if } x < 0 \\ \sum_{x=0}^{[x]} \binom{x+r-1}{r-1} p^r (1-p)^x, & \text{if } x \ge 0 \end{cases} \qquad \ldots (12.34)$$

Let us do one example.

Example 2: A biased coin which has the probability of getting a head as 1/4 and a tail as 3/4 is tossed till, we get the third head. If X counts the number of failures before the third head, then obtain the PMF and CDF of X and also plot the PMF and CDF of X.

Solution: If we define getting a head as a success and getting a tail as a failure then the random variable X which counts the number of failures before the third head follows a negative binomial distribution with parameters r = 3, p = 1/4. So, the PMF of X is given by putting r = 3, p = 1/4 in (12.25)

$$p_X(x) = \mathcal{P}(X=x) = \begin{cases} \binom{x+3-1}{3-1} \left(\frac{1}{4}\right)^3 \left(1-\frac{1}{4}\right)^x, & x = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

Or

$$p_X(x) = \mathcal{P}(X = x) = \begin{cases} \binom{x+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^x, & x = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases} \quad \ldots (12.35)$$

and the CDF of X is given by

$$F_X(x) = \mathcal{P}(X \le x) = \begin{cases} 0, & \text{if } x < 0 \\ \sum_{x=0}^{[x]} \binom{x+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^x, & \text{if } x \ge 0 \end{cases} \quad \ldots (12.36)$$

Probabilities for X = 0, 1, 2, 3, …, 10 are given by

$$\left.\begin{aligned}
\mathcal{P}(X=0) &= \binom{0+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^0 = 0.015625 \\
\mathcal{P}(X=1) &= \binom{1+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^1 = 0.03515625 \\
\mathcal{P}(X=2) &= \binom{2+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^2 = 0.05273438 \\
\mathcal{P}(X=3) &= \binom{3+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^3 = 0.06591797 \\
\mathcal{P}(X=4) &= \binom{4+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^4 = 0.07415771 \\
\mathcal{P}(X=5) &= \binom{5+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^5 = 0.0778656 \\
\mathcal{P}(X=6) &= \binom{6+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^6 = 0.0778656 \\
\mathcal{P}(X=7) &= \binom{7+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^7 = 0.07508469 \\
\mathcal{P}(X=8) &= \binom{8+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^8 = 0.07039189 \\
\mathcal{P}(X=9) &= \binom{9+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^9 = 0.06452590 \\
\mathcal{P}(X=10) &= \binom{10+2}{2}\left(\frac{1}{64}\right)\left(\frac{3}{4}\right)^{10} = 0.05807331
\end{aligned}\right\} \quad \ldots (12.37)$$

Now, using (12.36) CDF of X is given by

$$F_X(x) = \mathcal{P}(X \le x) = \begin{cases} 0, & \text{if } x < 0 \\ 0.015625, & \text{if } 0 \le x < 1 \\ 0.05078125, & \text{if } 1 \le x < 2 \\ 0.10351563, & \text{if } 2 \le x < 3 \\ 0.16943359, & \text{if } 3 \le x < 4 \\ 0.24359131, & \text{if } 4 \le x < 5 \\ 0.32145691, & \text{if } 5 \le x < 6 \\ 0.39932251, & \text{if } 6 \le x < 7 \\ 0.4744072, & \text{if } 7 \le x < 8 \\ 0.54479909, & \text{if } 8 \le x < 9 \\ 0.60932499, & \text{if } 9 \le x < 10 \\ 0.66739830, & \text{if } 10 \le x < 11 \\ \vdots & \end{cases} \quad \ldots (12.38)$$

PMF and CDF are plotted in Fig. 12.2 (a) and (b) respectively as follows.

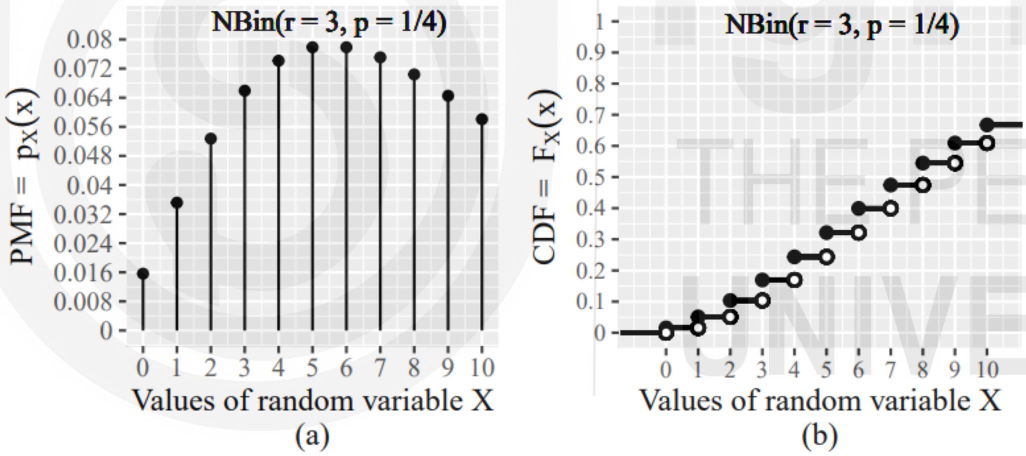


Fig. 12.2: Visualisation of (a) PMF (b) CDF of the NBin(3, 1/4) discussed in Example 2

12.5 MGF AND OTHER SUMMARY MEASURES OF NEGATIVE BINOMIAL DISTRIBUTION

In the previous section, you have studied PMF and CDF of negative binomial distribution. In this section, we want to obtain MGF, mean and variance of negative binomial distribution.

Calculation of MGF

Suppose we perform Bernoulli trials with probability of success p till, we get r^{th} success. Let $X_i, i = 1, 2, 3, \ldots, r$ be the random variable which counts the number of failures between i – 1 and the i^{th} success. So, each $X_i \sim \text{Geom}(p), i = 1, 2, 3, \ldots, r.$ Using (12.16), (12.17) and (12.20), we have

$$M_{X_i}(t)=\frac{p}{1-(1-p)e^t},\ t<\ln\left(\frac{1}{1-p}\right),\ i=1,2,3,\ldots,r \qquad \ldots (12.39)$$

$$E\left(X_i\right)=\frac{1-p}{p},\ i=1,2,3,\ldots,r \qquad \ldots (12.40)$$

$$V\left(X_i\right)=\frac{1-p}{p^2},\ i=1,2,3,\ldots,r \qquad \ldots (12.41)$$

Let $X=X_1+X_2+X_3+\ldots+X_r$ … (12.42)

$$M_X(t)=M_{X_1+X_2+X_3+\ldots+X_r}(t)=M_{X_1}(t)M_{X_2}(t)\ldots M_{X_r}(t)\ [\text{Using (7.94) of this course}]$$

$$=\prod_{i=1}^{r}M_{X_i}(t) \qquad [\because \text{ each } X_i \sim \text{Geom}(p),\ i=1,2,3,\ldots,r]$$

$$=\prod_{i=1}^{r}\frac{p}{1-(1-p)e^t},\ t\in\mathbb{R} \quad [\text{Using (12.39)}]$$

$$=\left(\frac{p}{1-(1-p)e^t}\right)^r$$

$$M_X(t)=\left(\frac{p}{1-(1-p)e^t}\right)^r \qquad \ldots (12.43)$$

Calculation of Mean

Applying expectation on both sides of (12.42), we have

$$E(X)=E\left(X_1+X_2+X_3+\ldots+X_r\right)$$

$$=E\left(X_1\right)+E\left(X_2\right)+E\left(X_3\right)+\ldots+E\left(X_r\right) \quad \left[\begin{array}{l}\text{Using addition theorem of}\\ \text{expectation refer to (7.29)}\end{array}\right]$$

$$=\underbrace{\frac{p}{1-p}+\frac{p}{1-p}+\frac{p}{1-p}+\ldots+\frac{p}{1-p}}_{r-\text{times}} \qquad [\text{Using (12.40)}]$$

$$=\frac{rp}{1-p}$$

Hence, mean of $X=$ expected value of $X=E(X)=\dfrac{rp}{1-p}$ … (12.44)

Calculation of Variance

Applying variance operator V(.) on both sides of (12.42), we have

$$V(X)=V\left(X_1+X_2+X_3+\ldots+X_r\right)$$

$$=V\left(X_1\right)+V\left(X_2\right)+V\left(X_3\right)+\ldots+V\left(X_r\right) \quad \left[\begin{array}{l}\text{Here, } X_1, X_2, X_3, \ldots, X_r \text{ are}\\ \text{independent. So, using (7.74)}\end{array}\right]$$

$$=\underbrace{\frac{p}{(1-p)^2}+\frac{p}{(1-p)^2}+\frac{p}{(1-p)^2}+\ldots+\frac{p}{(1-p)^2}}_{r-\text{times}} \quad [\text{Using (12.41)}]$$

$$=\frac{rp}{(1-p)^2}$$

Hence, variance of $X=V(X)=\dfrac{rp}{(1-p)^2}$ … (12.45)

We know that standard deviation of X is positive square root of variance of X.

Hence, standard deviation $= \text{SD}(X) = \sqrt{\text{Variance of } X} = \dfrac{\sqrt{rp}}{(1-p)}$ … (12.46)

Similarly, we can obtain other measures of our interest. But keeping applied nature of the programme in view, we are not focusing on proof of each measure. Some commonly used summary measures of negative binomial distribution are shown in Table 12.2 given as follows.

Table 12.2: Summary measures of negative binomial distribution

Name of measure	Formula	Name of measure	Formula
Mean	$\dfrac{rp}{1-p}$	MGF	$\left(\dfrac{p}{1-(1-p)e^t}\right)^r$
Variance	$\dfrac{rp}{(1-p)^2}$	Skewness	$\dfrac{2-p}{\sqrt{r(1-p)}}$
Standard deviation	$\dfrac{\sqrt{rp}}{(1-p)}$	Kurtosis	$\dfrac{6}{r}+\dfrac{p^2}{r(1-p)}$

12.6 APPLICATIONS AND ANALYSIS OF GEOMETRIC AND NEGATIVE BINOMIAL DISTRIBUTIONS

In this section, we will apply geometric and negative binomial distributions to solve some problems where assumptions of these distributions make sense. We will also do some analysis of these two distributions.

Example 3: Probability that Kavita hits a target is 0.7. Find the probability that Kavita hits the target in her third attempt. Assume that all the trials are independent.

Solution: Here it is given that trials are independent which means probability of hitting the target does not change trial to trial and, we are given that constant probability of hitting the target in each trial is p = 0.7. We are interested in getting the probability of the first success. So, it is the situation of geometric distribution. Let X be the random variable which counts the number of failures before the first success. Hence, X ~ Geom(0.7). Thus, required probability is given by

$$\mathcal{P}(X=2) = (1-0.7)^2(0.7) \qquad \left[\because \mathcal{P}(X=x) = \begin{cases}(1-p)^x p, & x = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise}\end{cases}\right]$$

$$= 0.063$$

Example 4: **Application in solving a game problem**: Arnav and Abhishek usually play table Tanis. On the basis of past experience, it is known that the probability that Arnav beats Abhishek is 0.55. Assume that each game is independent of the other. One day they decide to play till one of them wins five games. Find the probability that result of eighth game decide the winner.

Solution: Let X be the random variable which counts the number of failures before fifth success of Arnav while Y be the random variable which counts the number of failures before fifth success of Abhishek. Since each game is independent of the other so the probability that Arnav beats Abhishek is constant in each game and therefore, X ~ NBin(r = 5, p = 0.55). Similarly, Y ~ NBin(r = 5, p = 0.45). Now, result of the eighth game will decide winner of the game if either 8th game is the 5th success of Arnav or 8th game is the 5th success of Abhishek. So, required probability is given by

$$\mathcal{P}(X=3)+\mathcal{P}(Y=3)=\binom{3+5-1}{5-1}(0.55)^5(1-0.55)^3+\binom{3+5-1}{5-1}(0.45)^5(1-0.45)^3$$

$$,\left[\because \text{ if } Z\sim \text{NBin(r, p), then } \mathcal{P}(Z=z)=\begin{cases}\binom{x+r-1}{r-1}p^r(1-p)^z, & \text{if } x=0,1,2,3,\ldots\\ 0, & \text{otherwise}\end{cases}\right]$$

$$\Rightarrow \mathcal{P}(X=3)+\mathcal{P}(Y=3)=\binom{7}{4}(0.55)^5(0.45)^3+\binom{7}{4}(0.45)^5(0.55)^3$$

$$=0.2679693 \qquad \ldots (12.47)$$

Example 5: **Application in solving collector's problem**: There is a sequence of n toys. Suppose each packet of a particular food item contains one of these n toys. Assume that each toy is equally likely to be packed in the packet of this particular food item. Suppose you are interested in collecting complete sequence of toys. What is the expected value of the number of packets of food items you should buy.

Solution: Obviously the first packet that you will buy will contain one of the n toys. So, after purchasing one packet of this particular food you have one toy in your hand out of the total n toys. When you will buy the second food packet then it may contain the toy which you already have with you or you can get one among the remaining n – 1 toys. So, probability of getting additional toy is $\frac{n-1}{n}$ and probability of getting the existing toy is $\frac{1}{n}$. Let X_1 be the random variable which counts the number of failures before the first success when you buy second food item onward. So, $X_1 \sim \text{Geom}\left(p_1=\frac{n-1}{n}\right)$. We know that expected value of geometric distribution with parameter p_1 is $1/p_1$.. So, in the present case expected value of X_1 is given by $E(X_1)=\frac{1}{p_1}=\frac{n}{n-1}$. … (12.48)

So, expected number of food packets that you should buy to get two different toys is $1+\frac{n}{n-1}$. … (12.49)

After having two different toys in your hand, when you will go for buying the more food items then either the new packet of food may contain one from the two toys which you already have with you or it may contain the one among the remaining n – 2 toys. So, probability of getting additional toy is $\frac{n-2}{n}$ and probability of getting the one from the two existing toys is $\frac{2}{n}$. Let X_2 be the random variable which counts the number of failures before the first success where counting starts after getting two toys of different types. So,

$X_2 \sim \text{Geom}\left(p_2 = \frac{n-2}{n}\right)$. Expected value of X_2 is given by $E(X_2) = \frac{1}{p_2} = \frac{n}{n-2}$. … (12.50)

So, expected number of food packets that you should buy to get three different toys is $1 + \frac{n}{n-1} + \frac{n}{n-2}$. … (12.51)

Continuing in this fashion, the expected number of food packets that you should buy to get all the toys of the sequence of n toys is given by

$$1 + \frac{n}{n-1} + \frac{n}{n-2} + \frac{n}{n-3} + \ldots + \frac{n}{n-(n-2)} + \frac{n}{n-(n-1)}$$

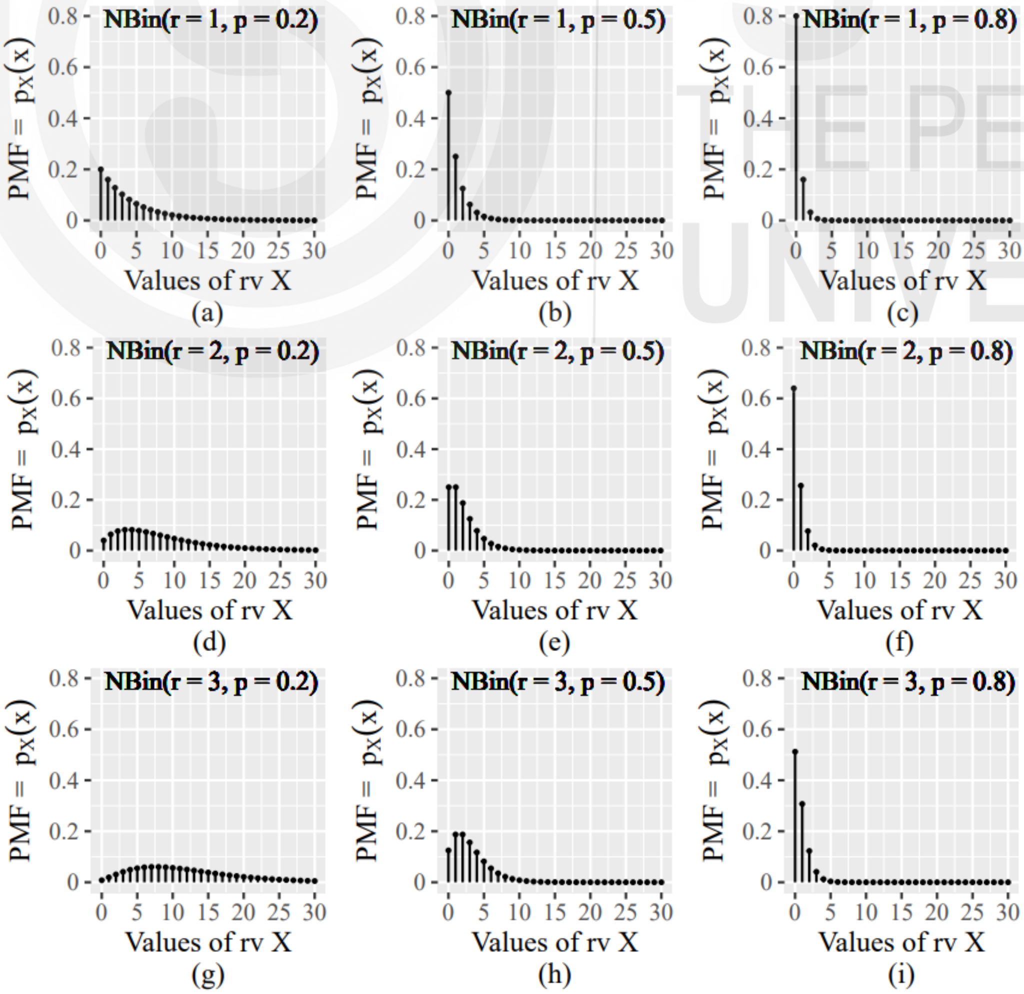
$$= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \frac{n}{n-3} + \ldots + \frac{n}{2} + \frac{n}{1}$$

$$= \frac{n}{1} + \frac{n}{2} + \frac{n}{3} + \frac{n}{4} + \ldots + \frac{n}{n-2} + \frac{n}{n-1} + \frac{n}{n} \quad \text{[Writing terms in reverse order]}$$

$$= n\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots + \frac{1}{n-2} + \frac{1}{n-1} + \frac{1}{n}\right) \quad \ldots (12.52)$$

Example 6: **Analysis of PMF of NBin(r, p) as r and p vary**: Plot PMF of negative binomial for different combinations of values of r and p where r = 1, 2, 3, 4, 5 and p = 0.2, 0.5, 0.8.

Solution: PMF of negative binomial distribution when p = 1 to 5 and p = 0.2, 0.5, 0.8 are shown in Fig. 12.3 (a) to (o) given as follows.



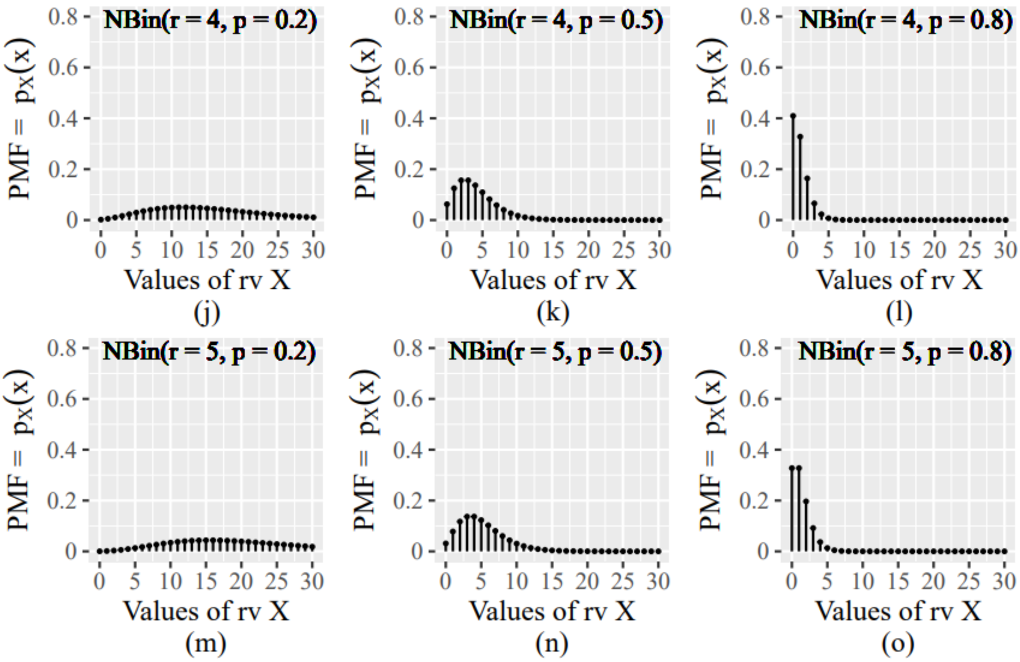


Fig. 12.3: Visualisation of PMF of negative binomial distributions when (a) r = 1, p = 0.2 (b) r = 1, p = 0.5 (c) r = 1, p = 0.8 (d) r = 2, p = 0.2 (e) r = 2, p = 0.5 (f) r = 2, p = 0.8 (g) r = 3, p = 0.2 (h) r = 3, p = 0.5 (i) r = 3, p = 0.8 (j) r = 4, p = 0.2 (k) r = 4, p = 0.5 (l) r = 4, p = 0.8 (m) r = 5, p = 0.2 (n) r = 5, p = 0.5 (o) r = 5, p = 0.8

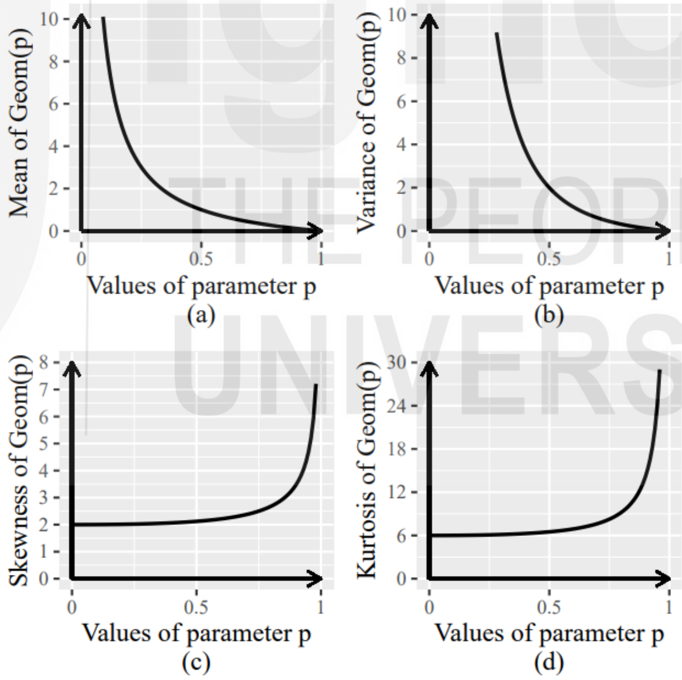


Fig. 12.4: Visualisation of summary measures (a) mean (b) variance (c) skewness (d) kurtosis of geometric distribution against values of p

Example 7: **Analysis of mean, variance, skewness and kurtosis of Geom(p) as p varies**: Plot mean, variance, skewness and kurtosis of geometric distribution as p varies. Also, specify some important observations noted in the graphical analysis of these summary measures.

Solution: If X ~ Geom(p), then, we know that mean, variance, skewness and kurtosis of geometric distribution are given by

$$\text{mean} = \frac{1-p}{p},\ \text{variance} = \frac{1-p}{p^2},\ \text{skewness} = \frac{2-p}{\sqrt{1-p}},\ \text{and kurtosis} = 6 + \frac{p^2}{1-p}.$$

Mean, variance, skewness and kurtosis of geometric distribution as p varies are shown in Fig. 12.4 (a) to (d).

Some important observations from the graphical analysis of these summary measures are mentioned as follows.

- As p increases mean and variance of the distribution both continuously decrease and tend to 0 if p tends to 1 while they tend to infinity as p tends to 0. At p = 1/2 value of mean is 1 while value of variance is 2. Variance increases more rapidly than mean.
- Minimum value of skewness is 2 when p = 0 and it remains almost 2 till $p \le 1/2$. But as p crosses 1/2 it starts to increase. But p always remains greater than equal to 2 so geometric distribution is always positively highly skewed distribution to know why you may refer to (9.78) to (9.81).
- Minimum value of kurtosis is 6 when p = 0 and it remains almost 6 till $p \le 1/2$. But as p crosses 1/2 it starts to increase. But p always remains greater than equal to 6 so geometric distribution is always leptokurtic refer (9.90).

Example 8: **Analysis of mean, variance, skewness and kurtosis of NBin(r, p) as r and p vary**: Plot mean, variance, skewness and kurtosis of negative binomial distribution as p varies for at least three different values of r. Also, specify some important observations noted in the graphical analysis of these summary measures. After getting the idea from the graphs comment on the skewness and kurtosis of the distribution.

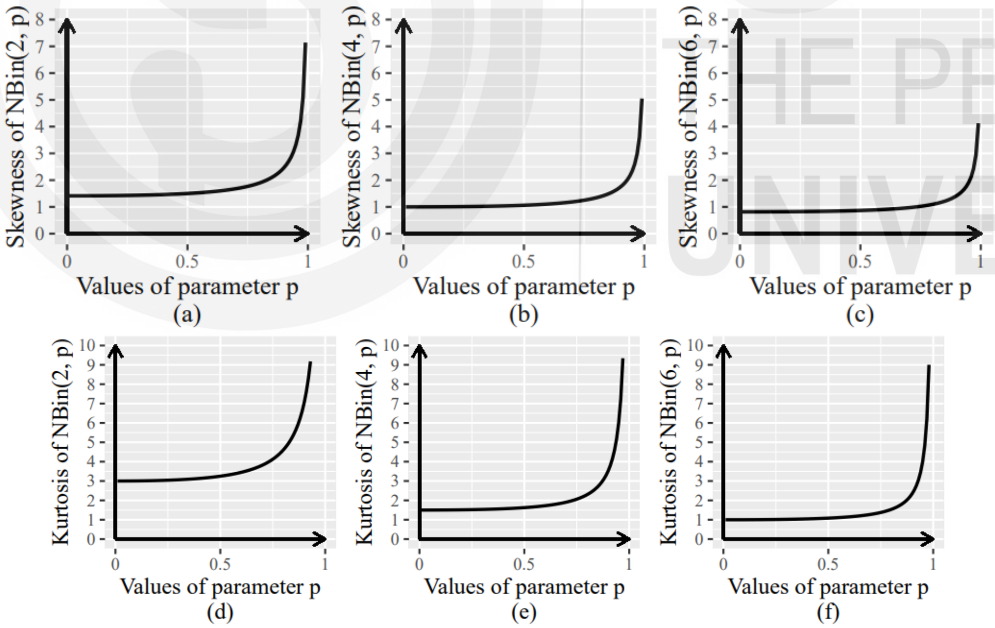


Fig. 12.5: Visualisation of summary measures (a) skewness when r = 2 (b) skewness when r = 4 (c) skewness when r = 6 (d) kurtosis when r = 2 (e) kurtosis when r = 4 (f) kurtosis when r = 6

Solution: If X ~ NBin(r, p), then, we know that mean, variance, skewness and kurtosis of negative binomial distribution are given by

$$\text{mean} = \frac{r(1-p)}{p}, \text{ variance} = \frac{r(1-p)}{p^2}, \text{ skewness} = \frac{2-p}{\sqrt{r(1-p)}}, \text{ kurtosis} = \frac{6}{r} + \frac{p^2}{r(1-p)}$$

We know that when r = 1 then negative binomial distribution becomes geometric distribution and, we have already done this type of analysis for geometric distribution in Example 7. So, let us consider three values of r as 2, 4 and 6. But mean and variance are just r times the mean and variance of the geometric distribution so there is no need of their graphical analysis. But r is sitting in denominator in the formulae of skewness and kurtosis so as r increases, they will decrease. So, let us analysis their graphical behaviour as p varies for r = 2, 4, 6 and are shown in Fig. 12.5 (a) to (f).

Some important observations from the graphical analysis of these summary measures are mentioned as follows.

- As p increases mean and variance of the distribution both continuously decrease and tend to 0 if p tends to 1 while they tend to infinity as p tends to 0. At p = 1/2 value of mean is 1 while value of variance is 2. Variance increases more rapidly then mean.
- Minimum value of skewness is 2 when p = 0 and it remains almost 2 till $p \leq 1/2.$ But as p crosses 1/2 it starts to increase. But p always remains greater than equal to 2 so geometric distribution is always positively highly skewed distribution to know why you may refer to (9.78) to (9.81).
- Minimum value of kurtosis is 6 when p = 0 and it remains almost 6 till $p \leq 1/2.$ But as p crosses 1/2 it starts to increase. But p always remains greater than equal to 6 so geometric distribution is always leptokurtic refer (9.90).

Observation: From Fig. 12.3 (a) to (o), we note that as value of r increases or value of p decreases PMF of negative binomial distribution spread on horizontal axis towards right.

Now, you can try the following two Self-Assessment Questions.

SAQ 1

Arnav and Abhishek usually play table Tanis. On the basis of past experience, it is known that the probability that Arnav beats Abhishek is 0.55. Assume that each game is independent of the other. One day they decide to play till one of them wins five games. If we are given that 8^{th} game has decided the winner then find the probability that Arnav is the winner.

SAQ 2

A person has 10 keys in his hand and he really does not know which one is the right key of the lock he wants to open. Out of these 10 keys there is only one key which can open the desired lock. Assume that he tries keys one by one randomly by selecting one key out of the 10 keys in his hand. Find the expected number of try he has to do.

12.7 SUMMARY

A brief summary of what we have covered in this unit is given as follows:

- **Definition and PMF of Geometric Distribution**: If we perform Bernoulli trials till, we get the first success and the random variable X counts the number of failures before the first success then the PMF of X is given by

$$p_X(x) = \mathcal{P}\left(X = x\right) = \begin{cases} (1-p)^x\, p, & x = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

This is the PMF of geometric distribution with parameter p and is denoted by writing $X \sim \text{Geom}(p)$.

- **CDF of Geometric Distribution**: Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}\left(X \le x\right) = \begin{cases} 0, & \text{if } x < 0 \\ 1-(1-p)^{[x]+1}, & \text{if } x \ge 0 \end{cases}$$

- **Summary measures of Poisson distribution**

Name of measure	Formula	Name of measure	Formula
Mean	$\frac{1-p}{p}$	MGF	$\frac{p}{1-(1-p)e^t}$
Variance	$\frac{1-p}{p^2}$	Skewness	$\frac{2-p}{\sqrt{1-p}}$
Standard deviation	$\frac{\sqrt{1-p}}{p}$	Kurtosis	$6+\frac{p^2}{1-p}$

- **Definition of Negative Binomial Distribution**: If we perform Bernoulli trial where probability of success is p and probability of failure is q = 1 – p till, we get r^{th} success then PMF of the random variable X which counts the number of failures before the r^{th} success is given by

$$p_X(x) = \mathcal{P}\left(X = x\right) = \begin{cases} \binom{x+r-1}{r-1} p^r (1-p)^x, & \text{if } x = 0, 1, 2, 3, \ldots \\ 0, & \text{otherwise} \end{cases}$$

- **CDF of Negative Binomial Distribution**: Let x be any fixed real number then CDF of the random variable X is given by

$$F_X(x) = \mathcal{P}\left(X \le x\right) = \begin{cases} 0, & \text{if } x < 0 \\ \sum_{x=0}^{[x]} \binom{x+r-1}{r-1} p^r (1-p)^x, & \text{if } x \ge 0 \end{cases}$$

- **Summary measures of Hypergeometric distribution**

Name of measure	Formula	Name of measure	Formula
Mean	$\frac{rp}{1-p}$	MGF	$\left(\frac{p}{1-(1-p)e^t}\right)^r$
Variance	$\frac{rp}{(1-p)^2}$	Skewness	$\frac{2-p}{\sqrt{r(1-p)}}$
Standard deviation	$\frac{\sqrt{rp}}{(1-p)}$	Kurtosis	$\frac{6}{r}+\frac{p^2}{r(1-p)}$

12.8 TERMINAL QUESTIONS

1. Arnav and Abhishek usually play table Tanis. On the basis of past experience, it is known that the probability that Arnav beats Abhishek is

0.55. Assume that each game is independent of the other. One day they decide to play till one of them wins five games. Find the probability that eventually Arnav is the winner on that day.

2. Suppose you have a 20 faces fair die. Using result of Example 5 or otherwise find the expected of throws to get every face of the die as the outcome.

3. **Lack of Memory Property or Memoryless Property of Geometric Distribution**: State and prove that geometric distribution has memoryless property. **Statement**: If X ~ Geom(p) then prove that $\mathcal{P}\left(X \geq j+k \mid X \geq j\right)=\mathcal{P}\left(X \geq j\right)$.

4. Why the name negative binomial distribution is given to the distribution of the random variable which counts the number of failures before the r^{th} success.

12.9 SOLUTIONS/ANSWERS

Self-Assessment Questions (SAQs)

1. Let the random variables X and Y be the same as explained in Example 4 of this unit. Now, 8^{th} game has decided the winner it means winner may be Arnav or Abhishek. So, using conditional probability required probability is given by

$$\mathcal{P}\left(X=3 \mid X=3 \text{ or } Y=3\right)=\frac{\mathcal{P}\left(\left(X=3\right)\cap\left(X=3 \text{ or } Y=3\right)\right)}{\mathcal{P}\left(X=3 \text{ or } Y=3\right)}$$

$$\left[\because \mathcal{P}\left(A \mid B\right)=\frac{\mathcal{P}\left(A\cap B\right)}{\mathcal{P}\left(B\right)}\right]$$

$$=\frac{\mathcal{P}\left(X=3\right)}{\mathcal{P}\left(X=3\right)+\mathcal{P}\left(X=3\right)} \qquad \text{... (12.53)}$$

$$\left[\begin{array}{l}\because \text{Events } X=3 \text{ and } Y=3\\ \text{are mutually exclusive}\end{array}\right]$$

we have already obtained value of denominator in Example 4 which is 0.2679693. So, let us first obtain the value of numerator.

$$\mathcal{P}\left(X=3\right)=\binom{3+5-1}{5-1}(0.55)^5(1-0.55)^3$$

$$\left[\because \mathcal{P}\left(X=x\right)=\begin{cases}\binom{x+r-1}{r-1}p^r(1-p)^x, & \text{if } x=0,1,2,3,\ldots\\ 0, & \text{otherwise}\end{cases}\right]$$

$$\Rightarrow \mathcal{P}\left(X=3\right)=\binom{7}{4}(0.55)^5(0.45)^3=0.1605163$$

Using values of numerator and denominator in (12.53), we get

$$\mathcal{P}\left(X=3 \mid X=3 \text{ or } Y=3\right)=\frac{0.1605163}{0.2679693}=0.59901.$$

2. Out of the 10 keys in his hand only one can open the lock so probability of opening the lock is 1/10. If we call opening the lock as success then probability of success is p = 1/10. Since each time he selects a key

randomly out of the 10 keys in his hand so probability of success in each trial is constant. Let the random variable X counts the number of failure before the first success. So, X ~ Geom(1/10). Hence, expected number of tries will be simply mean of the geometric distribution with parameter p = 1/10. But we know that expected value of mean of the geometric distribution is 1/p. Thus, expected number of tries he has to do is 1/(1/10) = 10.

Terminal Questions

1. As per the rules of the game to win, Arnav has to win 5 games before Abhishek wins 5 games. So, Arnav can be winner of that day if either his 5th game is his 5th success or 6th game is his 5th success or 7th game is his 5th or 8th game is his 5th success or 9th game is his 5th success. In other words, either Arnav has 0 failures before 5th success or 1 failure before 5th success or 2 failures before 5th success or 3 failures before 5th success or at the most 4 failures before 5th success. So, required probability is given by (recall that in this problem X counts the number of failures before the fifth success)

$$\mathcal{P}(X=0 \text{ or } 1 \text{ or } 2 \text{ or } 3 \text{ or } 4)=\mathcal{P}(X=0)+\mathcal{P}(X=1)+\mathcal{P}(X=2)+\mathcal{P}(X=3)+\mathcal{P}(X=4) \quad \ldots (12.54)$$

$$\left[\because \text{Events } X=0,\ X=1,\ X=2,\ X=3 \text{ and } X=4 \text{ are mutually exclusive}\right]$$

$$\Rightarrow \mathcal{P}(X=0 \text{ or } 1 \text{ or } 2 \text{ or } 3 \text{ or } 4)=\binom{0+5-1}{5-1}(0.55)^5(1-0.55)^0$$

$$+\binom{1+5-1}{5-1}(0.55)^5(1-0.55)^1+\binom{2+5-1}{5-1}(0.55)^5(1-0.55)^2$$

$$+\binom{3+5-1}{5-1}(0.55)^5(1-0.55)^3+\binom{4+5-1}{5-1}(0.55)^5(1-0.55)^4$$

$$\left[\because \mathcal{P}(X=x)=\begin{cases}\binom{x+r-1}{r-1}p^r(1-p)^x, & \text{if } x=0,1,2,3,\ldots\\ 0, & \text{otherwise}\end{cases}\right]$$

$$\Rightarrow \mathcal{P}(X=0 \text{ or } 1 \text{ or } 2 \text{ or } 3 \text{ or } 4)=\binom{4}{4}(0.55)^5(0.45)^0$$

$$+\binom{5}{4}(0.55)^5(0.45)^1+\binom{6}{4}(0.55)^5(0.45)^2$$

$$+\binom{7}{4}(0.55)^5(0.45)^3+\binom{8}{4}(0.55)^5(0.45)^4$$

$$=0.6214209 \qquad \left[\text{Using scientific calculator}\right]$$

In your lab exam you will obtain value of such probability using following R code. Screenshot with output is shown as follows.

```
> pnbinom(q = 4, size = 5, prob = 0.55)
[1] 0.6214209
```

2. Using result of Example 5 the expected number of throws to get every face of the die as the outcome is given by

$$20\left(\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\ldots+\frac{1}{18}+\frac{1}{19}+\frac{1}{20}\right)=71.95479\simeq 72.$$

3. We know that PMF of geometric distribution with parameter p is given by

$$p_X(x)=\mathcal{P}(X=x)=\begin{cases}(1-p)^x\,p, & x=0,1,2,3,\ldots\\ 0, & \text{otherwise}\end{cases}$$

We claim that $\mathcal{P}(X\geq j)=(1-p)^j$.

$$\begin{aligned}\mathcal{P}(X\geq j)&=\mathcal{P}(X\geq j)+\mathcal{P}(X\geq j+1)+\mathcal{P}(X\geq j+2)+\ldots\\ &=(1-p)^j p+(1-p)^{j+1}p+(1-p)^{j+2}p+\ldots\\ &=(1-p)^j p\left[1+(1-p)+(1-p)^2+\ldots\right]\\ &=(1-p)^j p\left[\frac{1}{1-(1-p)}\right]\quad\left[\because \text{sum of infinite GP } a+ar+ar^2+\ldots=\frac{a}{1-r},\ \text{provided } |r|<1\right]\\ &=\frac{(1-p)^j p}{p}=(1-p)^j\end{aligned}\qquad \ldots (12.55)$$

Now,

$$\begin{aligned}\mathcal{P}(X\geq j+k\,|\,X\geq j)&=\frac{\mathcal{P}((X\geq j+k)\cap(X\geq j))}{\mathcal{P}(X\geq k)}\quad\left[\because \mathcal{P}(A\,|\,B)=\frac{\mathcal{P}(A\cap B)}{\mathcal{P}(B)}\right]\\ &=\frac{\mathcal{P}(X\geq j+k)}{\mathcal{P}(X\geq k)}\\ &=\frac{(1-p)^{j+k}}{(1-p)^k}\quad[\text{Using (12.55)}]\\ &=(1-p)^k\end{aligned}$$

The above result reveals that the conditional probability of at least first $j+k$ trials are unsuccessful before the first success given that at least first j trials were unsuccessful, is the same as the probability that the first k trials were unsuccessful. So, the probability to get first success remains same if we start counting of k unsuccessful trials from anywhere provided all the trials preceding to it are unsuccessful, i.e., the future does not depend on past, it depends only on the present. So, the geometric distribution forgets the preceding trials and hence this property is given the name “forgetfulness property” or “Memoryless property” or “lack of memory” property.

4. Using (12.32) in (12.25) PMF of negative binomial distribution is given by

$$p_X(x)=\mathcal{P}(X=x)=\begin{cases}(-1)^x\dbinom{-r}{x}p^r(1-p)^x, & \text{if } x=0,1,2,3,\ldots\\ 0, & \text{otherwise}\end{cases}$$

Let us consider

$$p_X(x) = \mathcal{P}\left(X = x\right) = \left(-1\right)^x \binom{-r}{x} p^r\,(q)^x, \text{ if } x = 0, 1, 2, 3, \ldots \left[\because 1 - p = q\right]$$

$$= \binom{-r}{x} p^r\,(-q)^x, \text{ if } x = 0, 1, 2, 3, \ldots$$

$$= \binom{-r}{x} (-q)^x (1)^{-r-x} p^r, \text{ if } x = 0, 1, 2, 3, \ldots \qquad \left[\because (1)^{-r-x} = 1\right]$$

So, probabilities for X = 0, 1, 2, 3, 4, … are successive terms of the binomial expansion given as follows.

$$\left(1 + (-q)\right)^{-r} p^r \qquad \left[\because (a+b)^n = \binom{n}{0} a^n b^0 + \binom{n}{1} a^{x-1} b + \binom{n}{2} a^x b^{n-x} + \ldots\right]$$

$$= \left(1 + (-q)\right)^{-r} \left(\frac{1}{p}\right)^{-r}$$

$$= \left(\frac{1-q}{p}\right)^{-r}$$

$$= \left(\frac{1}{p} + \frac{-q}{p}\right)^{-r}$$

Which is a binomial expansion with negative index. That is why it is known as negative binomial distribution.