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

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

where $0 \le p \le 1$ and $\binom{n}{x} = \dfrac{\lfloor n}{\lfloor x \lfloor n-x} = \text{binomial coefficient}$ … (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:

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

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

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 $\underbrace{\text{SSFFFSFSFFSSSF}\ldots\text{FS}}_{\text{n-times}}$ 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,

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

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

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

$$\mathcal{P}(S)\mathcal{P}(F)\mathcal{P}(S)\mathcal{P}(F)\mathcal{P}(F)\ldots\mathcal{P}(F)\mathcal{P}(S) \begin{bmatrix}\text{If E and F are independent} \\ \text{events, then} \\ \mathcal{P}(E\cap F) = \mathcal{P}(E)\mathcal{P}(F)\end{bmatrix}$$

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

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

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

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,

$\mathcal{P}\left(X = x\right) = \binom{n}{x} p^x (1-p)^{n-x}$, 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 $\Omega = \left\{\omega_1\omega_2\omega_3 \ldots \omega_n : \text{ each } \omega_i = \text{success or failure}\right\}$ contains all possible 2^n sequences of success and failures 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 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 $0 \le p \le 1$

So, PMF of binomial distribution is given by

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

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

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

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

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

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

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.

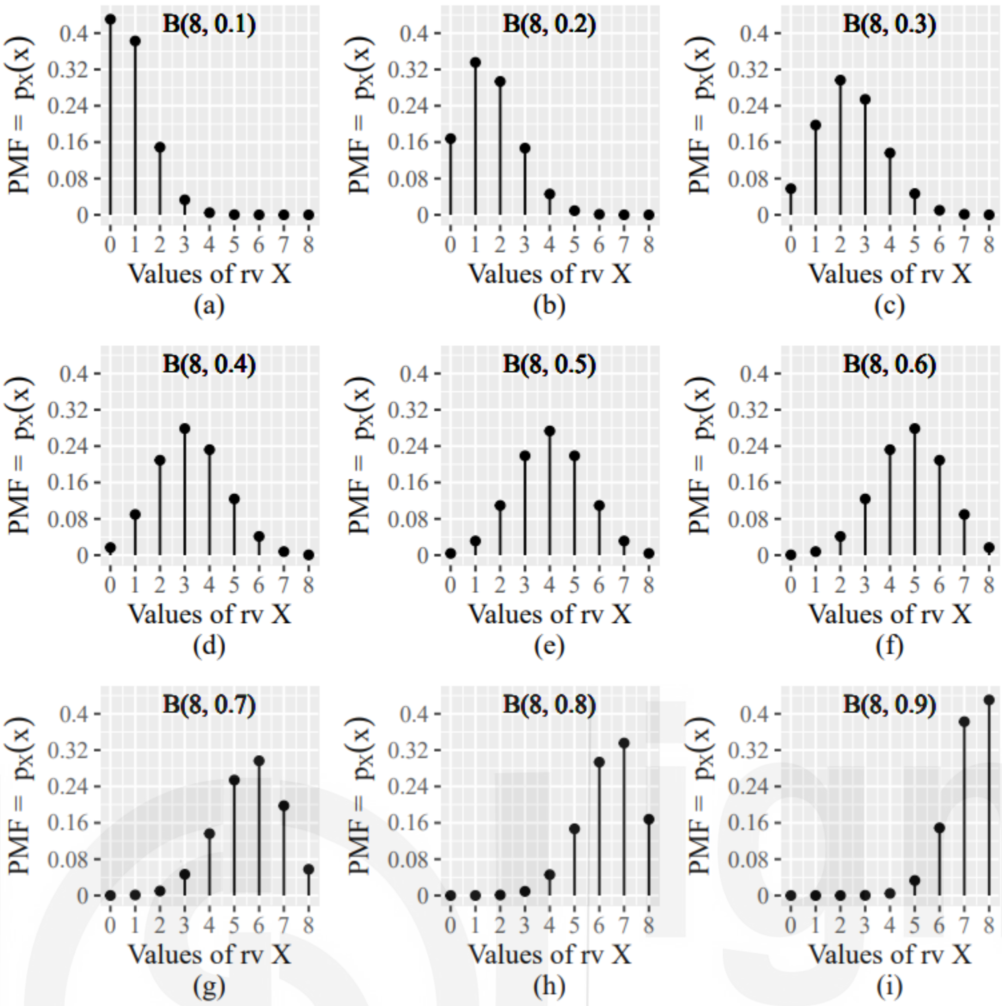

Fig. 10.1: Visualisation of PMF of binomial distributions where n = 8 and p is (a) 0.1 (b) 0.2 (c) 0.3 (d) 0.4 (e) 0.5 (f) 0.6 (g) 0.7 (h) 0.8 (i) 0.9

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:

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

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

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

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

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

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

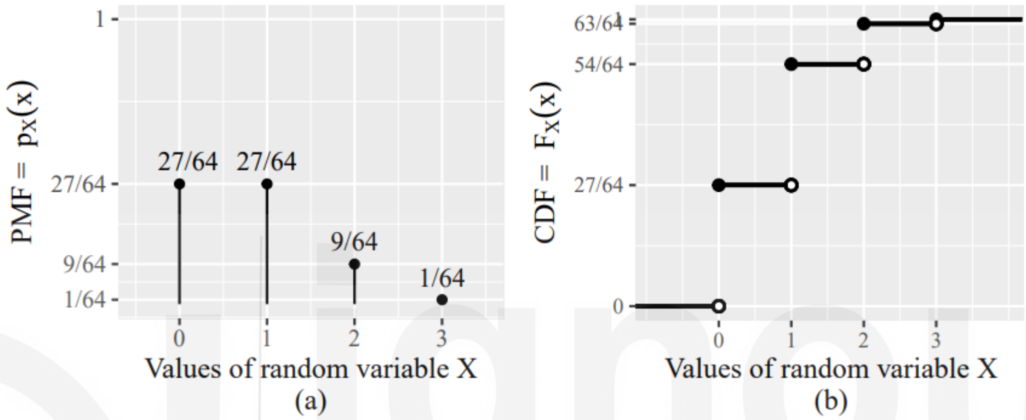

Fig. 10.2: Visualisation of (a) PMF (b) CDF of the B(3, 1/4) discussed in Example 2

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

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

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

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

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

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

$$= \left(q+pe^t\right)\left(q+pe^t\right)\left(q+pe^t\right)\ldots\left(q+pe^t\right) \qquad [\text{Using } (9.55)]$$

$$= \left(q+pe^t\right)^n \qquad \ldots (10.28)$$

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

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

$$= E\left(I_1\right)+E\left(I_2\right)+E\left(I_3\right)+\ldots+E\left(I_n\right) \qquad \left[\begin{array}{l}\text{Using addition theorem of}\\ \text{expectation refer to (7.30)}\end{array}\right]$$

$$= p+p+p+\ldots+p \qquad [\text{Using } (10.26)]$$

$$= np \qquad \ldots (10.29)$$

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

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

$$= V\left(I_1\right)+V\left(I_2\right)+V\left(I_3\right)+\ldots+V\left(I_n\right) \qquad [\text{Using } (7.74)]$$

$$= p(1-p)+p(1-p)+p(1-p)+\ldots+p(1-p) \qquad [\text{Using } (10.26)]$$

$$= np(1-p) \qquad \ldots (10.30)$$

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

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

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

or $x \le p(n+1)-1$ … (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	$(q+pe^t)^n$
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 $(a_1 + a_2 + a_3 + \ldots + a_m)^n$ is given by

$$(a_1 + a_2 + a_3 + \ldots + a_m)^n = \sum_{n_1, n_2 \ldots n_m} \frac{\underline{|n}}{\underline{|n_1}\,\underline{|n_2}\,\underline{|n_3} \cdots \underline{|n_m}} a_1^{n_1} a_2^{n_2} a_3^{n_3} \ldots a_m^{n_m} \quad \ldots (10.36)$$

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.

$$(a+b)^7 = \binom{7}{0}a^7b^0 + \binom{7}{1}a^6b + \binom{7}{2}a^5b^2 + \binom{7}{3}a^4b^3 + \binom{7}{4}a^3b^4 + \binom{7}{5}a^2b^5 + \binom{7}{6}ab^6 + \binom{7}{7}a^0b^7$$

$$= a^7b^0 + 7a^6b + 21a^5b^2 + 35a^4b^3 + 35a^3b^4 + 21a^2b^5 + 7ab^6 + a^0b^7 \quad (10.38)$$

Also, note that

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

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

$$\underbrace{aaaaabb + aaaaabb + aaaaabb + \ldots + aaaaabb}_{\binom{7}{5}\text{- times} = 21\text{ times}}$$

$$= \underbrace{a^5b^2 + a^5b^2 + a^5b^2 + \ldots + a^5b^2}_{21\text{ times}} = 21a^5b^2 \quad \ldots (10.40)$$

Now, we apply this argument to understand the multinomial expansion given by (10.36). To obtain the general term $\frac{\underline{|n}}{\underline{|n_1}\,\underline{|n_2}\,\underline{|n_3} \cdots \underline{|n_m}} a_1^{n_1} a_2^{n_2} a_3^{n_3} \ldots a_m^{n_m}$ 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

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

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

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

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

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

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

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

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 $\mathcal{P}\left(E_i\right)=p_i,\ i=1, 2, 3, \ldots, k.$ 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 $\mathcal{P}\left(E_i\right)=p_i,\ i=1, 2, 3, \ldots, k.$ 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 $\mathcal{P}(E_i) = p_i,\ i = 1, 2, 3, \ldots, k$ should remain the same in all trials. That is $\mathcal{P}(E_i) = p_i,\ i = 1, 2, 3, \ldots, k$ 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

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

So, each random variable $X_i, 1 \le i \le k$ 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

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

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

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

… (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

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

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

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

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

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

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

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

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

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

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

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 $X_i X_j \ \forall \ i, j = 1, 2, 3, \ldots, k$

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

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

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

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

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

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

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

Now, using (7.67), we have

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

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

$$Cov(X_i, X_j) = n(n-1)p_i p_j - (np_i)(np_j)$$

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

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

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

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

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

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

… (10.62)

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

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

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

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

$$E\left((X - E(X))(X - E(X))'\right) = \begin{bmatrix} np_1(1-p_1) & -np_1p_2 & \cdots & -np_1p_k \\ -np_2p_1 & np_2(1-p_2) & \cdots & -np_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -np_kp_1 & -np_kp_2 & \cdots & np_k(1-p_k) \end{bmatrix} \text{… (10.65)}$$

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

Solution: Comparing $\text{multinom}(20, 1/10, 3/10, 6/10)$ with $\text{multinom}(n, p_1, p_2, p_3)$, 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 $X = (X_1, X_2, X_3)$ 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

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

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

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

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

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

- **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

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

- 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

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

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

$$= 0.02152828$$

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).

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

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

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

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

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

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

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

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