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).

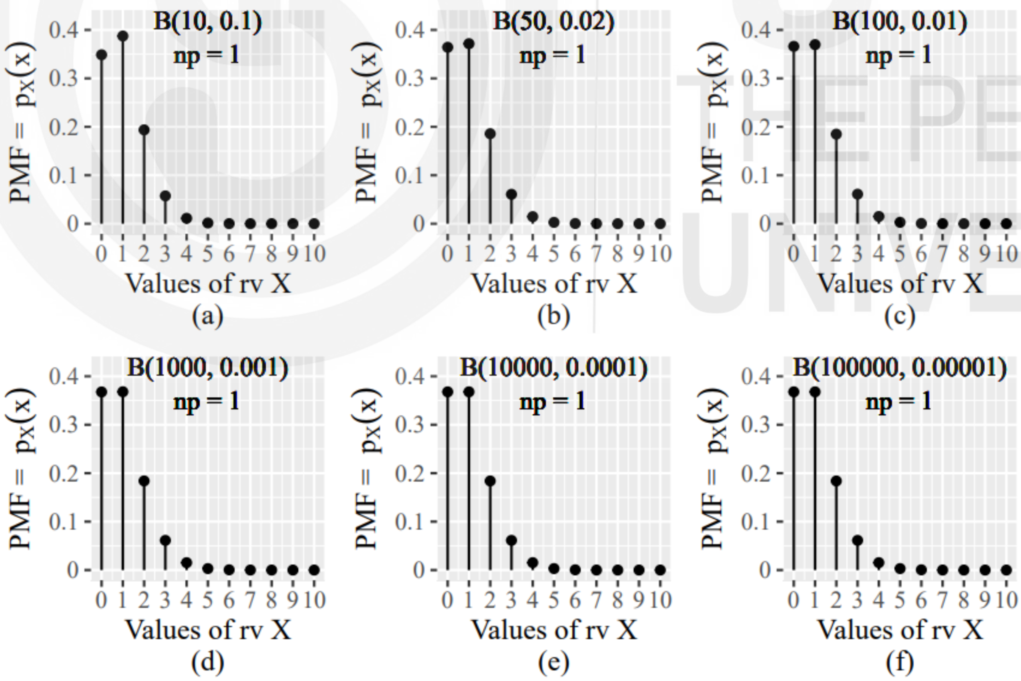

Fig. 11.1: Visualisation of PMF of binomial distributions where np = 1 but n and p vary (a) n = 10, p = 0.1 (b) n = 50, p = 0.02 (c) n = 100, p = 0.01 (d) n = 1000, p = 0.001 (e) n = 10000, p = 0.0001 (f) n = 100000, p = 0.00001

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).

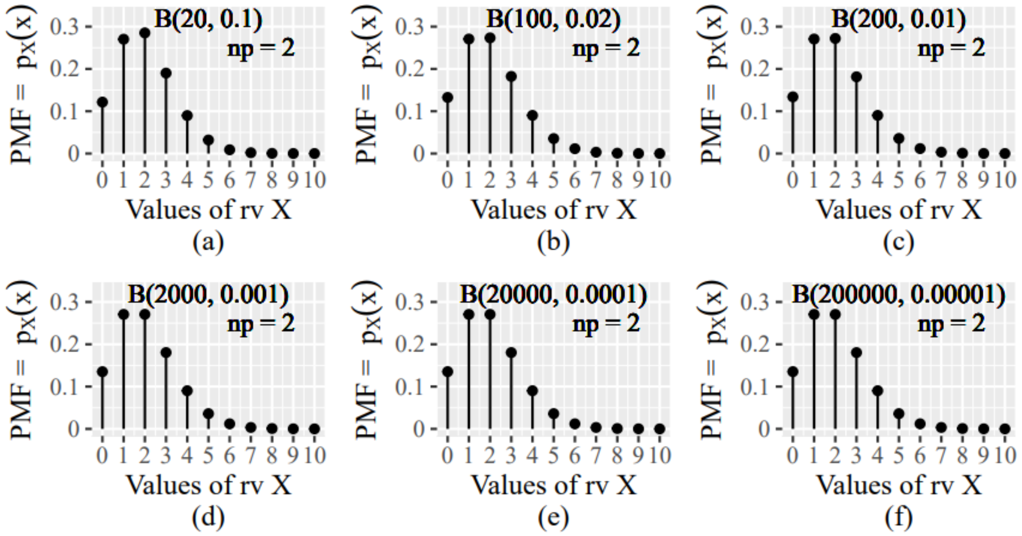

Fig. 11.2: Visualisation of PMF of binomial distributions where np = 2 but n and p vary (a) n = 20, p = 0.1 (b) n = 100, p = 0.02 (c) n = 200, p = 0.01 (d) n = 2000, p = 0.001 (e) n = 20000, p = 0.0001 (f) n = 200000, p = 0.00001

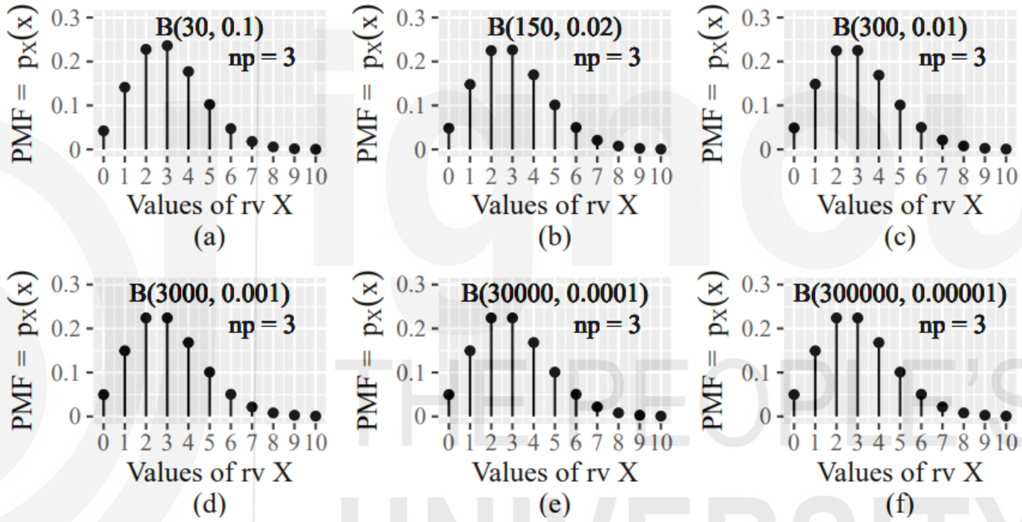

Fig. 11.3: Visualisation of PMF of binomial distributions where np = 3 but n and p vary (a) n = 30, p = 0.1 (b) n = 150, p = 0.02 (c) n = 300, p = 0.01 (d) n = 3000, p = 0.001 (e) n = 30000, p = 0.0001 (f) n = 300000, p = 0.00001

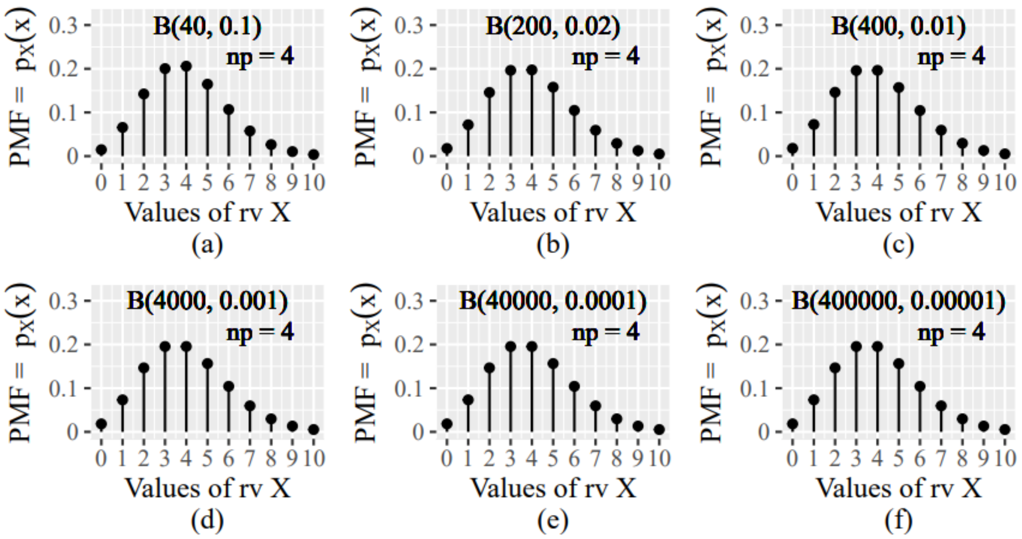

Fig. 11.4: Visualisation of PMF of binomial distributions where np = 4 but n and p vary (a) n = 40, p = 0.1 (b) n = 200, p = 0.02 (c) n = 400, p = 0.001 (d) n = 4000, p = 0.001 (e) n = 40000, p = 0.0001 (f) n = 400000, p = 0.00001

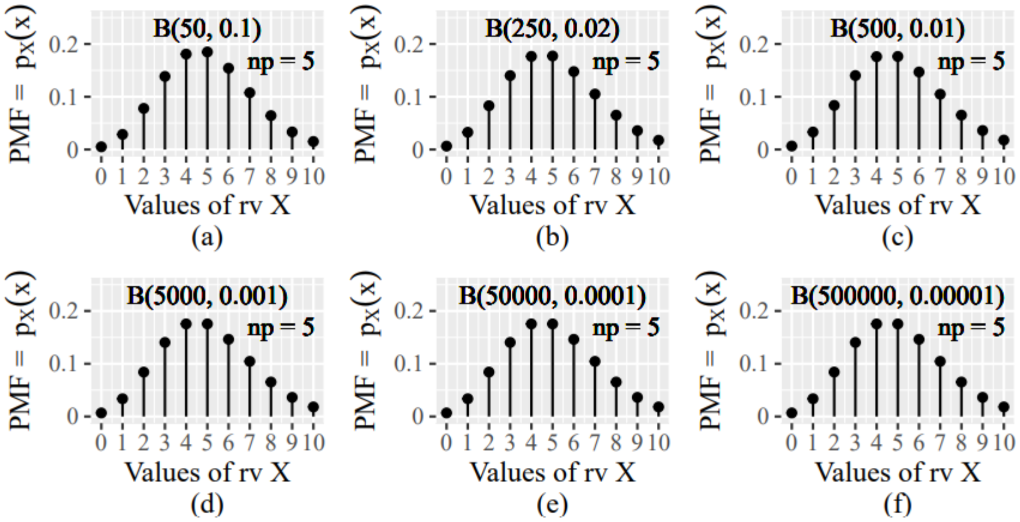

Fig. 11.5: Visualisation of PMF of binomial distributions where np = 5 but n and p vary (a) n = 50, p = 0.1 (b) n = 250, p = 0.02 (c) n = 500, p = 0.01 (d) n = 5000, p = 0.001 (e) n = 50000, p = 0.0001 (f) n = 500000, p = 0.00001

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.

$$n \to \infty \quad \text{… (11.7)}$$

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

p is small … (11.9)

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

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

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

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

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

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

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 λ. Or we read it as X follows a Poisson distribution with parameter λ. … (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

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

(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

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

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

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

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.

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

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

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

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

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

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

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

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

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

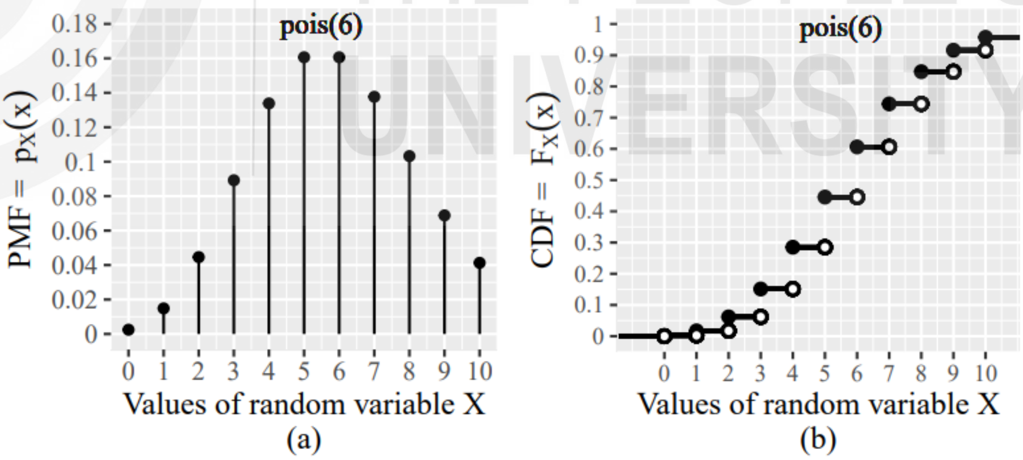

Fig. 11.6: Visualisation of (a) PMF (b) CDF of the Pois(6) discussed in Example 1

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

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

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

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

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

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

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

Calculation for Mean

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

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

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

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

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

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

Calculation for Variance and Standard Deviation

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}} = \frac{\dfrac{\underline{|G}}{\underline{|g}\,\underline{|G-g}}\;\dfrac{\underline{|B}}{\underline{|b}\,\underline{|B-b}}}{\dfrac{\underline{|N}}{\underline{|n}\,\underline{|N-n}}} = \frac{\underline{|G}\,\underline{|B}\,\underline{|n}\,\underline{|N-n}}{\underline{|g}\,\underline{|b}\,\underline{|G-g}\,\underline{|B-b}\,\underline{|N}} = \frac{\underline{|n}}{\underline{|g}\,\underline{|b}} \times \frac{\underline{|G}}{\underline{|G-g}} \times \frac{\underline{|B}}{\underline{|B-b}} \times \frac{\underline{|N-n}}{\underline{|N}}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

… (11.44)

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

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

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

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

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

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

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

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

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

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

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

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

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

Hence, mean of X = expected value of $X = E(X) = \dfrac{nG}{(G+B)}$ $\qquad \ldots (11.47)$

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 $X = V(X) = \frac{nGB(G+B-n)}{(G+B)^2(G+B-1)}$ … (11.50)

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

Hence, standard deviation $= SD(X) = \sqrt{\text{Variance of } X} = \sqrt{\frac{nGB(G+B-n)}{(G+B)^2(G+B-1)}}$

… (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	$\frac{nG}{(G+B)}$	Standard deviation	$\sqrt{\frac{nGB(G+B-n)}{(G+B)^2(G+B-1)}}$
Variance	$\frac{nGB(G+B-n)}{(G+B)^2(G+B-1)}$		

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 $X \sim \text{Pois}(\lambda = 25/3).$ 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 $\mathcal{P}\left(X+Y<4\right)$.

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

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

  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

$$F_X(x)=\mathcal{P}\left(X\le x\right)=\begin{cases}0, & \text{if } x<0 \\ \sum\limits_{k=0}^{[x]}\dfrac{e^{-\lambda}\lambda^k}{\underline{|k}}, & \text{if } 0\le x<\infty \\ 1, & \text{if } x\to\infty\end{cases}$$

- **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

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

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

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

- **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

$$\mathcal{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{\lfloor x}; \ x = 0, 1, 2, 3, \ldots$$

$$\therefore \ \mathcal{P}(X = 1) = \mathcal{P}(X = 2) \Rightarrow \frac{e^{-\lambda}\lambda^1}{\lfloor 1} = \frac{e^{-\lambda}\lambda^2}{\lfloor 2} \Rightarrow 2\lambda = \lambda^2 \Rightarrow \lambda^2 - 2\lambda = 0$$

$$\Rightarrow \lambda(\lambda - 2) = 0 \Rightarrow \lambda = 0, 2.$$

   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

$$\mathcal{P}(X = 3) = \frac{e^{-\lambda}\lambda^3}{\lfloor 3} = \frac{e^{-2}(2)^3}{6} = 0.180447. \quad \text{[Using scientific calculator]}$$