UNIT 9

UNIFORM AND BERNOULLI DISTRIBUTIONS

Structure

9.1 INTRODUCTION

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

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

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

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

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

Expected Learning Outcomes

After completing this unit, you should be able to:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

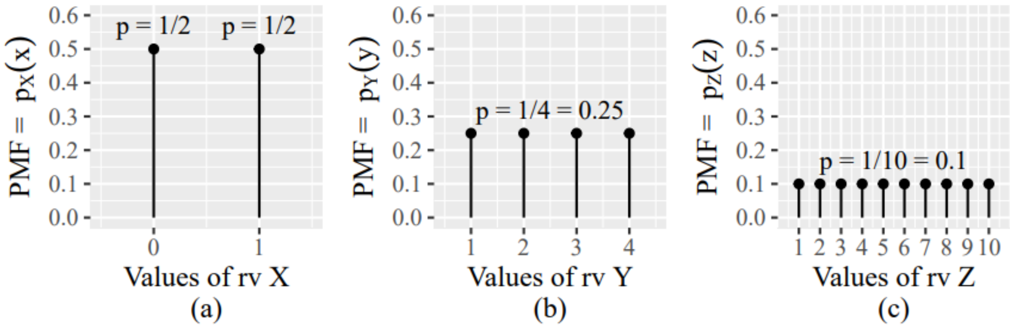

Fig. 9.1: Visualisation of PMF's of random variables (a) X (b) Y (c) Z

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

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

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

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

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

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

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

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

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

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

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

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

Now, we define discrete uniform probability distribution as follows.

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

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

where $b - a + 1 = N$ … (9.17)

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

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

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

So,

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

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

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

$\vdots$

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

So,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\sum_{k=a}^{b} \frac{1}{b - a + 1} = (b - a + 1)\frac{1}{b - a + 1} \quad \left[\because \sum_{k=1}^{n} a = na \text{ if } a \text{ is independent of } k\right]$$
$$= 1 \qquad \ldots (9.27)$$

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

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

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

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

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

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

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

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

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

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

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

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

Now, we define CDF of discrete uniform distribution.

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

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

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

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

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

Let us do one example.

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

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

PMF of the random variable X is given by

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

CDF of the random variable X is given by

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

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

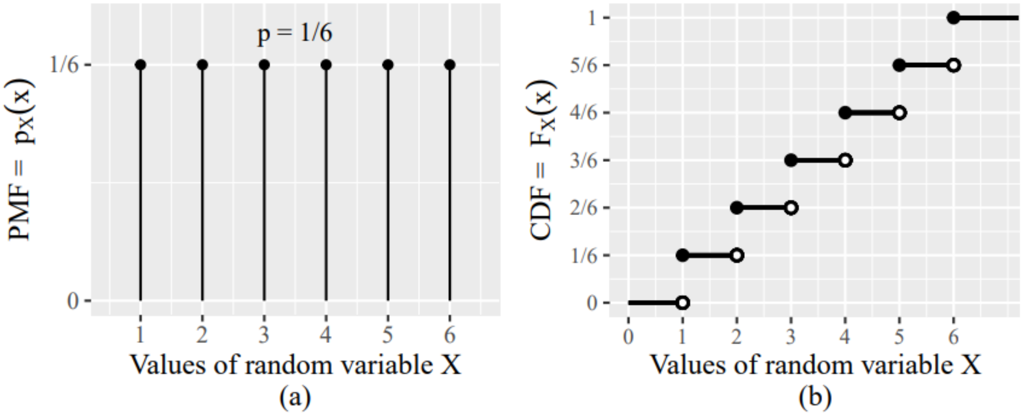

Fig. 9.2: Visualisation of (a) PMF of rv X (b) CDF of rv X of Example 1

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

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

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

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

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

9.3 MGF AND OTHER SUMMARY MEASURES OF DISCRETE UNIFORM DISTRIBUTION

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Putting x = y + a – 1

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table 9.4: Summary measures of discrete uniform distribution

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

9.4 STORY, DEFINITION, PMF AND CDF OF BERNOULLI DISTRIBUTION

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

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

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

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

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

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

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

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

Now, we define Bernoulli distribution as follows.

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

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

where $0 \le p \le 1$ … (9.45)

So, PMF of Bernoulli random variable X is given by

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

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

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

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

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

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

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

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

But $p + (1 - p) = 1$. … (9.51)

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

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

Now, we define CDF of Bernoulli distribution.

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

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

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

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

Let us do one example.

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

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

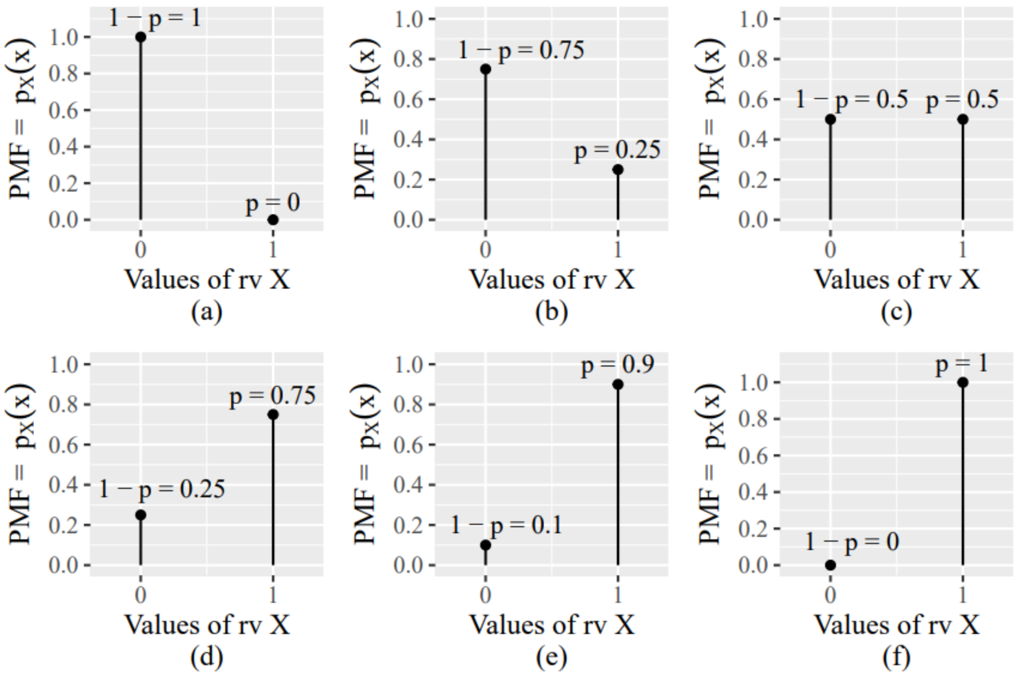

Fig. 9.3: Visualisation of PMF of Bernoulli distribution when p is (a) 0 (b) 0.25 (c) 0.5 (d) 0.75 (e) 0.9 (f) 1

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

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

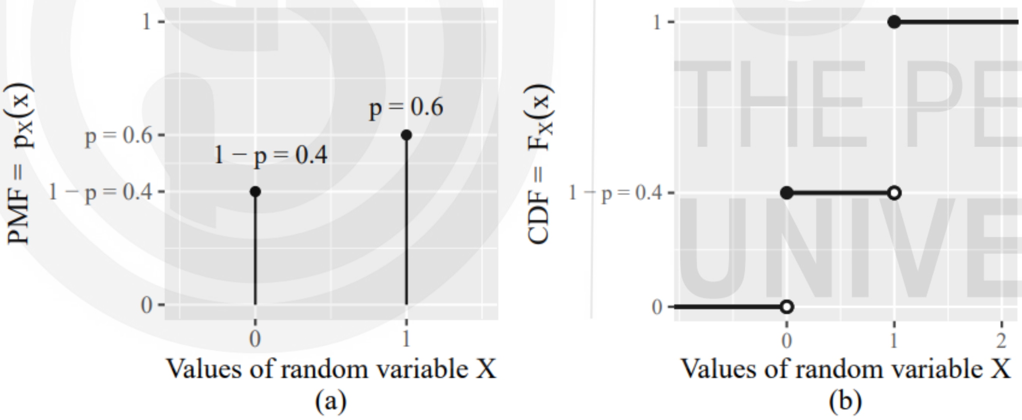

Fig. 9.4: Visualisation of (a) PMF (b) CDF of Bernoulli rv X when p = 0.6

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

9.5 MGF AND OTHER SUMMARY MEASURES OF BERNOULLI DISTRIBUTION

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

Calculation of MGF

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

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

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

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

where $q = 1 - p$

Calculation of Some Summary Measures

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

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

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

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

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

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

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

Hence,

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

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

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

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

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

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

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

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

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

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

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

Calculation for Mode

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

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

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

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

Calculation for Median

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

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

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

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

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

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

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

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

Table 9.5: Summary measures of Bernoulli distribution

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

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

9.6 APPLICATIONS AND ANALYSIS OF DISCRETE UNIFORM AND BERNOULLI DISTRIBUTIONS

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

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

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

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

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

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

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

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

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

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

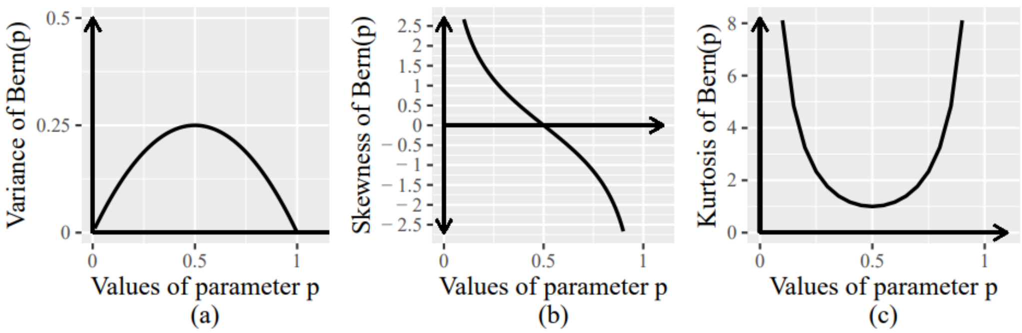

Fig. 9.5: Visualisation of (a) variance (b) skewness (c) kurtosis with respect to values of p of Bernoulli distribution with parameter p

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

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

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

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

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

  … (9.75)

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

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

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

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

… (9.78)

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

… (9.79)

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

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

… (9.81)

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

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

  … (9.82)

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

  … (9.83)

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

- So, Bernoulli distribution will be: approximately symmetric if

$$\frac{17-\sqrt{17}}{34} \le p \le \frac{17+\sqrt{17}}{34} \quad \left[\begin{array}{l} \because \dfrac{1-2p}{\sqrt{p(1-p)}} = \dfrac{1}{2} \Rightarrow 1+4p^2-4p = \dfrac{p-p^2}{4} \\ \Rightarrow 17p^2-17p+4=0 \Rightarrow p = \dfrac{-b \pm \sqrt{b^2-4ac}}{2a} = \dfrac{17 \pm \sqrt{17}}{34} \end{array}\right] \quad \text{… (9.85)}$$

moderately skewed if

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

If we want to write in decimal form then

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

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

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

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

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

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

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

- Distribution of Bern(p) will be mesokurtic if

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{aligned} E\left(\sum_{k=1}^{1920} X_k\right) &= \sum_{k=1}^{1920} E\left(X_k\right) && \left[\begin{array}{l}\text{Using addition theorem of}\\ \text{expectation refer to (7.30)}\end{array}\right] \\ &= \sum_{k=1}^{1920}\left(\frac{1}{60}\right) && \left[\text{Using (9.97)}\right] \\ &= 1920 \times \frac{1}{60} && \left[\because \sum_{k=1}^{n} a = na, \text{ if } a \text{ is independent of } k\right] \\ &= 32 && \ldots (9.98) \end{aligned}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hence, PMF of X_i is given by

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

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

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

SAQ 1

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

SAQ 2

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

9.7 SUMMARY

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

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

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

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

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

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

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

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

- **Summary measures of discrete uniform distribution**

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

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

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

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

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

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

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

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

- **Summary measures of Bernoulli distribution**

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

9.8 TERMINAL QUESTIONS

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

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

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

9.9 SOLUTIONS/ANSWERS

<u>Self-Assessment Questions (SAQs)</u>

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

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

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

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

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

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

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

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

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

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

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

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

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

Terminal Questions

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

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

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

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

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

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

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

Hence, required probability is given by

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

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

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

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

According to the question, we have

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

$$\Rightarrow a + b = 22 \qquad \text{... (9.103)}$$

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

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

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

$$\Rightarrow b - a = 4 \qquad \text{... (9.104)}$$

Adding (9.103) and (9.104), we get

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

Putting b = 13 in (9.103), we get

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

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

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

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

According to the question, we have

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

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

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

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

Reference

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

  DOI: 10.1080/00031305.2014.917055