UNIT 4

UNIVARIATE DISCRETE RANDOM VARIABLE

Structure

4.1 INTRODUCTION

In Unit 1 of this course, you have studied about sample space of a random experiment, events as subsets of a sample space and some rules to obtain the probability of an event of our interest. Consider a sample space of a random experiment of tossing three coins simultaneously. As usual, we denote sample space by Ω so $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$. To obtain the probability of an event of our interest related to this sample space, we have to directly deal with these 8 strings of H and T. But if somehow, we are able to convert these strings of H and T in numbers it would be a great relief to deal with numbers compare to strings of H and T. Random variable do this interesting and important job. How this job is done by a random variable is explained in Sec. 4.2. But to connect the concept of random variable with the foundation built in Units 2 and 3 another definition of random variable as a measurable function is discussed in Sec. 4.3. The concept of random variable can be used to develop another probability space known as induced probability space by the random variable and the same is discussed in Sec. 4.4. An important function which not only used to generalise the probability law from uniform to non-uniform but also plays the key role in moving from measure theory to probability theory is known as cumulative distribution function which is discussed in Sec. 4.5. After travelling a long journey of different ideas finally we will learn to apply these concepts in some examples in Sec. 4.6 where we will discuss discrete random variable and its probability mass function (PMF) and distribution function. A very special random variable known as indicator random variable is discussed in Sec. 4.7.

What we have discussed in this unit is summarised in Sec. 4.8. 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. 4.9 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. 4.10.

In the next unit, you will study univariate continuous random variables.

Expected Learning Outcomes

After completing this unit, you should be able to:

- explain the concept of random variable and why it is important in the field of probability theory;
- explain what is induced probability space;
- define cumulative distribution function and its properties;
- apply discrete random variable in solving a variety of problems; and
- explain a very simple but useful random variable known as indicator random variable.

4.2 RANDOM VARIABLE

In Unit 1, we discussed the sample space of a random experiment. Let us write sample space of four random experiments.

Random Experiment 1: Sample space of the random experiment of tossing a coin once is given by $\Omega_1 = \{H, T\}$. … (4.1)

Random Experiment 2: Sample space of the random experiment of tossing a coin twice or two coins simultaneously is given by $\Omega_2 = \{HH, HT, TH, TT\}$. (4.2)

Random Experiment 3: Sample space of the random experiment of tossing a coin thrice or three coins simultaneously is given by
$\Omega_3 = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$. … (4.3)

Random Experiment 4: Sample space of the random experiment of tossing a coin 100 times or 100 coins simultaneously is given as follows.

Here, the first row is written corresponding to the single possibility of 100 heads or 0 tail, the second row is written corresponding to the $\binom{100}{1} = 100$ possibilities of 99 heads and 1 tail, the third row is written corresponding to the $\binom{100}{2} = 4950$ possibilities of 98 heads and 2 tails, …, the second last row is written corresponding to the $\binom{100}{1} = 100$ possibilities of 1 head and 99 tails, and the last row is written corresponding to the single possibility of 0 head or 100 tails.

$$
\begin{aligned}
\Omega_4 = \{&\underbrace{HHH\ldots H}_{100-\text{times}},\\
&\underbrace{HHH\ldots H}_{99-\text{times}}T,\ \underbrace{HHH\ldots H}_{98-\text{times}}TH,\ \underbrace{HHH\ldots H}_{97-\text{times}}THH,\ \ldots,\ T\underbrace{HHH\ldots H}_{99-\text{times}},\\
&\underbrace{HHH\ldots H}_{98-\text{times}}TT,\ \underbrace{HHH\ldots H}_{97-\text{times}}THT,\ \ldots,\ TT\underbrace{HHH\ldots H}_{98-\text{times}},\ \ldots,\\
&\vdots\\
&H\underbrace{TTT\ldots T}_{99-\text{times}},\ TH\underbrace{TTT\ldots T}_{98-\text{times}},\ TTH\underbrace{TTT\ldots T}_{97-\text{times}},\ \ldots,\ \underbrace{TTT\ldots T}_{99-\text{times}}H,\\
&\underbrace{TTT\ldots T}_{100-\text{times}}\}.
\end{aligned}
\qquad \ldots (4.4)
$$

Suppose you are owner of a factory where some particular product is produced. If G denotes that produced item is good and D denotes that produced item is defective, then obviously you will not be interested in the sequence of G's and D's like the sequence of H's and T's in these experiments. You will simply be interested in the number of good items or defective items. Probability theory introduces the idea of the random variable to get and analyse this type of information. Actually, a random variable associates a unique real number to each outcome of the random experiment. That is, a random variable is a function from sample space to some subset of the set of all real numbers refer to Fig. 4.1.

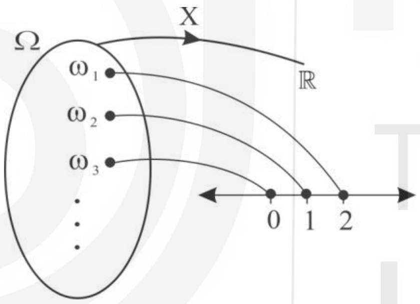

Fig. 4.1: Visualisation of the random variable as a function from sample space Ω of a random experiment to some subset B of the set of all real numbers or real line

For example, if X denotes the number of heads in the random experiment 3, then, we have

$$
\begin{aligned}
&X(HHH)=3,\ X(HHT)=2,\ X(HTH)=2,\ X(THH)=2,\\
&X(HTT)=1,\ X(THT)=1,\ X(TTH)=1,\ X(TTT)=0.
\end{aligned}
\qquad \ldots (4.5)
$$

A formal definition of a random variable is given as follows.

Random Variable: Let Ω be the sample space of a random experiment then a random variable X is a function from a sample space (Ω) of a random experiment to a subset of the set of all real numbers, i.e., $X:\Omega \to \mathbb{R}$. So, the domain of a random variable is the sample space of a random experiment and the range may be any subset of the set of all real numbers. Random variables are denoted by capital letters X, Y, Z, U, V, W, etc. … (4.6)

In Unit 1 of the course MST-011, you have learnt about pictorial presentation of a function. Random variable is also a function so it can also be presented pictorially. For example, if X denotes the number of heads, then random variables for random experiments 1 to 4 discussed earlier in this section can be visualised in pictorial form in Fig. 4.2 (a) to (d).

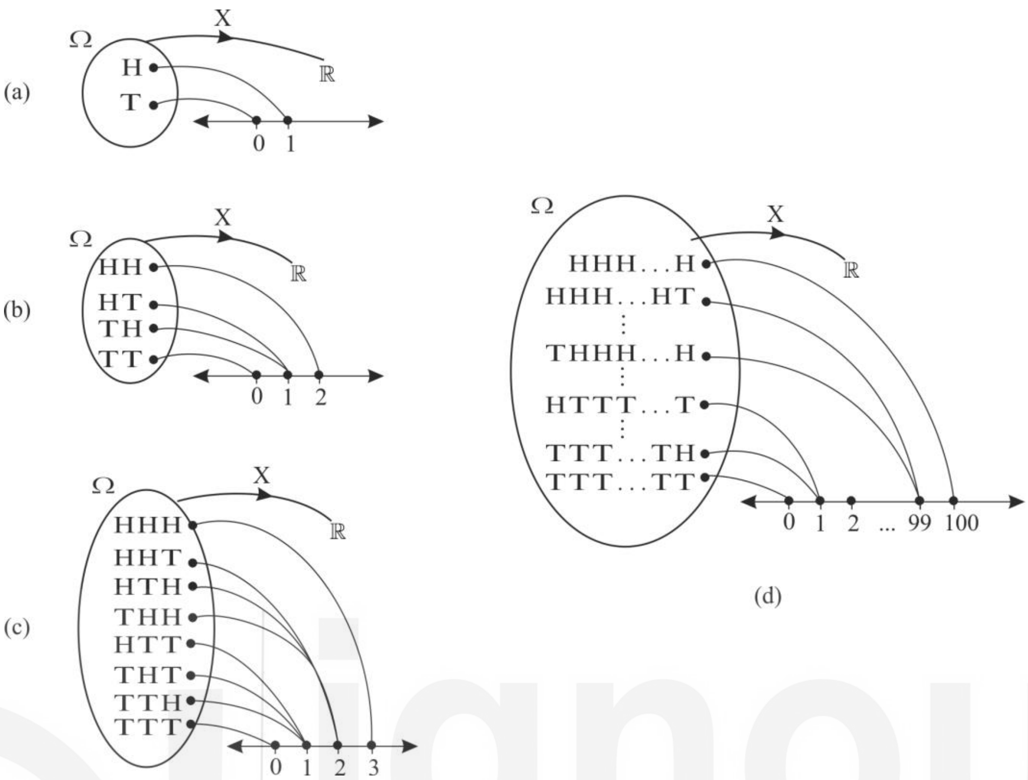

Fig. 4.2: Visualisation of the random variable as a function in pictorial form for the sample spaces of the random experiments (a) 1 (b) 2 (c) 3 (d) 4

In view of (4.5) and Fig. 4.2 (c), we see that 3 is **image** of HHH or HHH is **pre-image** of 3. We can also say that HHH is the **inverse image** of 3 under the random variable X and it is denoted by writing $X^{-1}(3) = HHH$. Similarly, we can write inverse images of other members 2, 1 and 0 under X which are shown in Fig. 4.2 (c) as follows.

$$\begin{aligned} X^{-1}(2) &= \{\omega \in \Omega : X(\omega) = 2\} = \{HHT, HTH, THH\} \\ X^{-1}(1) &= \{\omega \in \Omega : X(\omega) = 1\} = \{HTT, THT, TTH\} \qquad \ldots (4.7) \\ X^{-1}(0) &= \{\omega \in \Omega : X(\omega) = 0\} = \{TTT\} \end{aligned}$$

Now, let us discuss some examples which are not random variables. These examples will help you in understanding random variable in a better way.

Example 1: Suppose we have an unbiassed coin and we have coloured both faces of the coin. Suppose the head side face is done in red colour and the tail side face is done in black colour. If this coin is tossed then write sample space of this random experiment. Further, if X denotes the colour of the face that turned up. Will X be a random variable?

Solution: The sample space of this random experiment will be the same as shown in (4.1). If X denotes the colour of the face that turned up then X will associate red with outcome H and black with the outcome T. That is, we have

$$X(H) = \text{red}, \; X(T) = \text{black}.$$

Note that outcomes of X as a function are red and black which are not real numbers. So, X is not a random variable. In other words, range of the random variable X = {red, black} which is not a subset of the set of all real numbers. Hence, X is not a random variable. … (4.8)

However, if we define Y as Y(H) = 1 and Y(T) = 0, then Y will be a random variable because the range of Y is {0, 1} which is the subset of the set of real numbers.

Example 2: Consider the family of natural numbers which is denoted by $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$. Suppose X associates each natural number with its square. Will X be a random variable?

Solution: Here $X : \mathbb{N} \to \mathbb{N}$ is defined by

$X(n) = n^2, \; n \in \mathbb{N}$

Here range of X is {1, 4, 9, 14, …} which is a subset of the set of all real numbers. So, the range satisfies the requirement of a random variable. Now, we look at the domain of X. The domain of X is $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$ which may be the sample space of a random experiment, but in this example, it is not specified. So, the domain of X does not satisfy the requirement of the sample space of a random experiment. So, X is not a random variable. In other words, randomness is associated with the values of X only via randomness of the members of the sample space which is not mentioned here. So, for the randomness either the domain of X should be given as the sample space of a random experiment or direct probabilities are associated with the values of X to become a random variable. Here, none of the two is available. … (4.9)

So, remember the following two important points whenever you have to check whether a given function is a random variable or not.

- The domain of the function should be the sample space of a random experiment. It is the source of randomness which is conveyed via assigning probabilities to the values of X.
- The range of the function should be a subset of the set of all real numbers.

4.3 RANDOM VARIABLE AS A MEASURABLE FUNCTION

In Units 2 and 3, we have seen that measure theory is the mathematical foundation of the probability theory. So, it becomes important to understand the definition of random variable in the terminology of measure theory. But before that you have to understand what is a measurable function which is defined as follows.

Measurable and Borel Measurable Function: Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces (refer 2.29 to recall what is a measurable space) then a function $f : \Omega_1 \to \Omega_2$ is said to be a **measurable function** with respect to $\mathcal{F}_1, \mathcal{F}_2$ if $f^{-1}(B) \in \mathcal{F}_1$ for each $B \in \mathcal{F}_2$. In particular, if $\Omega_2 = \mathbb{R}, \; \mathcal{F}_2 = \mathcal{B}(\mathbb{R})$ then we say that f is **Borel measurable function**. … (4.10)

For example, let $(\Omega, \mathcal{F})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be two measurable spaces where $\mathcal{B}(\mathbb{R})$ is Borel σ-field on $\mathbb{R}$. You have studied characteristic function in Unit 3of the course MST-011, if you like you may refer (3.29). Consider the characteristic function χ_A of A where $A \in \mathcal{F}$, then by definition of the characteristic function, we have

$$\chi_A(a) = \begin{cases} 1, & a \in A \\ 0, & a \notin A \end{cases} \quad \ldots (4.11)$$

We claim that the characteristic function χ_A of A is a measurable function. Let $D \in \mathcal{B}(\mathbb{R})$ then four cases arise (i) both $0, 1 \in D$ (ii) $1 \in D$ but $0 \notin D$ (iii) $0 \in D$ but $1 \notin D$ (iv) $0 \notin D$ as well as $1 \notin D$. So,

$$\chi_A^{-1}(D) = \begin{cases} \Omega, & \text{if both } 0, 1 \in D \\ A, & \text{if } 1 \in D \text{ but } 0 \notin D \\ A^c, & \text{if } 0 \in D \text{ but } 1 \notin D \\ \phi, & \text{if } 0 \notin D \text{ as well as } 1 \notin D \end{cases} \quad \ldots (4.12)$$

Since ϕ, Ω belong to every σ-field on Ω. So, $\phi,\ \Omega \in \mathcal{F}$. Also, every σ-field is closed with respect to complement so $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$. Thus, all the inverse images Ω, A, A^c and ϕ of D under the characteristic function χ_A of A lies in $\mathcal{F}$. So, the characteristic function χ_A of A is measurable with respect to $\mathcal{F},\ \mathcal{B}(\mathbb{R})$. … (4.13)

We, now define random variable in terminology of measure theory as follows.

Random Variable in Terminology of Measure Theory: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be Borel measurable space. A function $X : \Omega \to \mathbb{R}$ is said to be a random variable with respect to $\mathcal{F}, \mathcal{B}(\mathbb{R})$ if $X^{-1}(B) \in \mathcal{F}$ for each $B \in \mathcal{B}(\mathbb{R})$. … (4.14)

By combining probability measure $\mathcal{P}$ and X^{-1} as defined in (4.14) we get an interesting probability measure known as induced probability measure which is discussed in the next section.

4.4 INDUCED PROBABILITY SPACE BY THE RANDOM VARIABLE ON THE REAL LINE

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be Borel measurable space. To convert Borel measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ into a probability space we need to define a probability measure on Borel σ-field $\mathcal{B}(\mathbb{R})$. Standard practice to define a probability measure on $\mathcal{B}(\mathbb{R})$ is by combining probability measure $\mathcal{P}$ of the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and a random variable $X : \Omega \to \mathbb{R}$

If $X : \Omega \to \mathbb{R}$ is a random variable then

$$X^{-1}(B) \in \mathcal{F} \text{ for each } B \in \mathcal{B}(\mathbb{R}). \quad \ldots (4.15)$$

Let us define a probability measure $\mathcal{P}_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$ by

$$\mathcal{P}_X(B) = \mathcal{P}\left(X^{-1}(B)\right),\ \ B \in \mathcal{B}(\mathbb{R}) \quad \ldots (4.16)$$

Why we defined $\mathcal{P}_X$ like this, you will get answer of this question in (4.29).

Since $X:\Omega \to \mathbb{R}$ is a random variable, so using (4.14), we have
$X^{-1}(B) \in \mathcal{F} \quad \forall\ B \in \mathcal{B}(\mathbb{R})$ … (4.17)

Also, $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space implies $\mathcal{P}:\mathcal{F} \to [0, 1]$ … (4.18)
is a probability measure. So, $\mathcal{P}$ will satisfy the following three conditions:

(i) $\mathcal{P}(\phi) = 0$ … (4.19)

(ii) $\mathcal{P}(\Omega) = 1$ … (4.20)

(iii) $\mathcal{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathcal{P}(A_k)$ … (4.21)

Due to (4.17) and (4.18), we have $\mathcal{P}\left(X^{-1}(B)\right) \in [0, 1]$ … (4.22)

Using (4.22) in (4.16), we have $\mathcal{P}_X(B) \in [0, 1] \quad \forall \quad B \in \mathcal{B}(\mathbb{R})$ … (4.23)

Due to (4.23), we can say that the map $\mathcal{P}_X$ is taking inputs from $\mathcal{B}(\mathbb{R})$ and is giving outputs in [0, 1]. So, to prove that $\mathcal{P}_X$ is a valid probability measure on $\mathcal{B}(\mathbb{R})$ we only need to prove that it satisfies (4.19) to (4.21). Let us prove these three requirements one at a time. To prove (4.19) for $\mathcal{P}_X,$ consider $\phi \in \mathcal{B}(\mathbb{R}).$ Due to (4.15), we have $X^{-1}(\phi) \in \mathcal{F}.$ We know that pre-image of an empty set is an empty set. So, $X^{-1}(\phi) = \phi.$ … (4.24)

$$\begin{aligned} \text{Now, } \mathcal{P}_X(\phi) &= \mathcal{P}\left(X^{-1}(\phi)\right) && [\text{Using } (4.16)] \\ &= \mathcal{P}(\phi) && [\text{Using } (4.24)] \\ &= 0 && [\text{Using } (4.19)] \end{aligned}$$

This proves (4.19) for $\mathcal{P}_X.$

To prove (4.20) for $\mathcal{P}_X,$ consider $\mathbb{R} \in \mathcal{B}(\mathbb{R}).$ Due to (4.15), we have $X^{-1}(\mathbb{R}) \in \mathcal{F}.$ We know that pre-image of whole set is Ω itself. So,

$X^{-1}(\mathbb{R}) = \Omega.$ … (4.25)

$$\begin{aligned} \text{Now, } \mathcal{P}_X(\mathbb{R}) &= \mathcal{P}\left(X^{-1}(\mathbb{R})\right) && [\text{Using } (4.16)] \\ &= \mathcal{P}(\Omega) && [\text{Using } (4.25)] \\ &= 1 && [\text{Using } (4.20)] \end{aligned}$$

This proves (4.20) for $\mathcal{P}_X.$

To prove (4.21) for $\mathcal{P}_X,$ consider $B_k \in \mathcal{B}(\mathbb{R}),\ k = 1, 2, 3, \ldots$ such that

$B_m \cap B_n = \phi \ \forall\ m \neq n$ … (4.26)

Due to (4.15), we have $X^{-1}(B_k) \in \mathcal{F},\ k = 1, 2, 3, \ldots$

Also,

$$\begin{aligned} X^{-1}(B_m) \cap X^{-1}(B_n) &= X^{-1}(B_m \cap B_n) && \left[\begin{array}{l}\because \text{pre-images respect intersection, so} \\ X^{-1}(A) \cap X^{-1}(B) = X^{-1}(A \cap B)\end{array}\right] \\ &= X^{-1}(\phi) && [\text{Using } (4.26)] \\ &= \phi && [\text{Using } (4.24)] \end{aligned}$$

Thus, $X^{-1}(B_k) \in \mathcal{F},\ k = 1, 2, 3, \ldots$ are disjoints.

$$\text{Now, } \mathcal{P}_X\left(\bigcup_{k=1}^{\infty} B_k\right) = \mathcal{P}\left(X^{-1}\left(\bigcup_{k=1}^{\infty} B_k\right)\right) \quad [\text{Using } (4.16)]$$

$$= \mathcal{P}\left(\bigcup_{k=1}^{\infty} X^{-1}(B_k)\right) \quad \left[\begin{array}{l}\because \text{pre-images respect countable union,} \\ \text{so } X^{-1}\left(\bigcup_{k=1}^{\infty} B_k\right) = \bigcup_{k=1}^{\infty} X^{-1}(B_k)\end{array}\right]$$

$$= \sum_{k=1}^{\infty} \mathcal{P}\left(X^{-1}(B_k)\right) \quad [\text{Using } (4.21)]$$

$$= \sum_{k=1}^{\infty} \mathcal{P}_X(B_k) \quad [\text{Using } (4.16)]$$

This proves (4.21) for $\mathcal{P}_X$.

Hence, $\mathcal{P}_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$ as defined by (4.16) is a valid probability measure on $\mathcal{B}(\mathbb{R})$ and so $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_X)$ is **probability space**. … (4.27)

The **probability measure** $\mathcal{P}_X$ is known as **probability measure induced by the random variable X on the real line**. The probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_X)$ with induced probability measure $\mathcal{P}_X$ is known as **induced probability space** by the random variable X on the real line. … (4.28)

Let us make two observations from the discussion of probability measure induced by the random variable on the real line.

(1) Let us visualise the whole process to see what is going on. Here our target was to define a probability measure on the Borel σ-field $\mathcal{B}(\mathbb{R})$. That is we were interested in getting a probability measure from $\mathcal{B}(\mathbb{R})$ to [0. 1]. To achieve this target, we started by considering B in $\mathcal{B}(\mathbb{R})$ refer Fig. 4.3. After that we use the definition of random variable and because of that pre-image $X^{-1}(B)$ of B will definitely belong to the σ-field $\mathcal{F}$ refer Fig. 4.3. Starting from $\mathcal{B}(\mathbb{R})$ we have reached in $\mathcal{F}$ with the help of the idea of pre-image of X. Now after reaching in $\mathcal{F}$, the good thing with $\mathcal{F}$ is we have available a probability measure $\mathcal{P}$ on $\mathcal{F}$ which assigns probability with each member of $\mathcal{F}$ and so it will assign probability to $X^{-1}(B)$ (the concept of assignment of probabilities to events has been discussed in detail in Unit 3). So, our requirement of getting a probability measure on $\mathcal{B}(\mathbb{R})$ can be converted into a reality if we combine the idea of pre-image $X^{-1}(B)$ of B and probability measure $\mathcal{P}$. That is why in (4.16) we defined a map $\mathcal{P}_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$ by

$\mathcal{P}_X(B) = \mathcal{P}\left(X^{-1}(B)\right),\ B \in \mathcal{B}(\mathbb{R}).$ … (4.29)

(2) Probability measure $\mathcal{P}_X$ plays a very crucial role in the definition of cumulative distribution function (CDF) and CDF plays a very crucial role indefining a probability distribution. CDF is discussed in the next section. But before that it becomes important to understand different ways of

writing (4.29). So, keep the following notations in mind all have the same meaning.

$$\mathcal{P}_X(B) = \mathcal{P}\left(X^{-1}(B)\right), \ B \in \mathcal{B}(\mathbb{R}) \quad \ldots (4.30)$$

$$\text{OR } \mathcal{P}_X(B) = \mathcal{P}\left(\{\omega \in \Omega : X(\omega) \in B\}\right), \ B \in \mathcal{B}(\mathbb{R}) \quad \ldots (4.31)$$

$$\text{OR } \mathcal{P}_X(B) = \mathcal{P}(X \in B), \ B \in \mathcal{B}(\mathbb{R}) \quad \ldots (4.32)$$

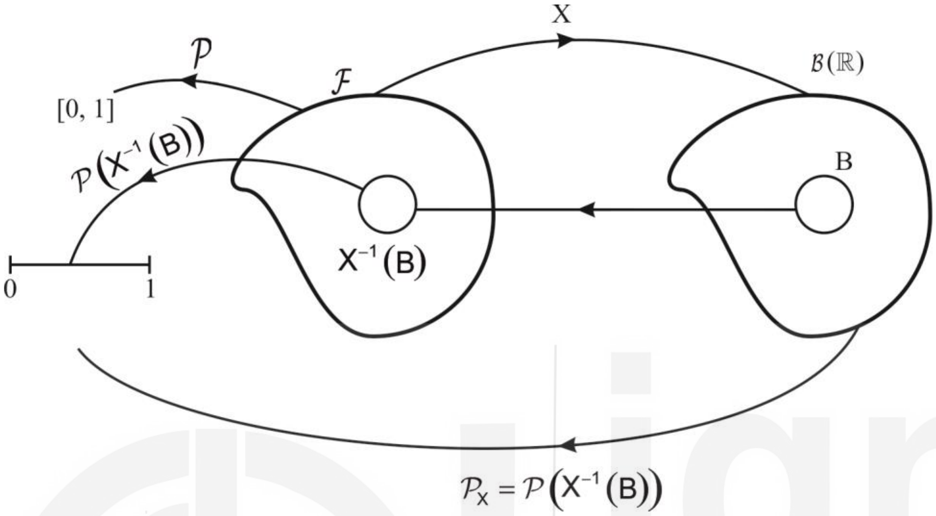

Fig. 4.3: Visualisation of the entire process of defining induced probability measure $\mathcal{P}_X$

4.5 CUMULATIVE DISTRIBUTION FUNCTION

In the previous section, we have defined a probability measure $\mathcal{P}_X$ on Borel σ-field $\mathcal{B}(\mathbb{R})$. So, the triple $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_X)$ is Borel probability space. It implies that the probability measure $\mathcal{P}_X$ assigns probability to every member of $\mathcal{B}(\mathbb{R})$. But we know that half open and half closed intervals of the form $(-\infty, x], x \in \mathbb{R}$ generate Borel σ-field $\mathcal{B}(\mathbb{R})$. That is $\sigma((-\infty, x] : x \in \mathbb{R}) = \mathcal{B}(\mathbb{R})$ and $(-\infty, x] \in \mathcal{B}(\mathbb{R})$ for all $x \in \mathbb{R}$. So, by definition of the probability measure $\mathcal{P}_X$ you may refer (4.16) or (4.30), we have

$$\mathcal{P}_X((-\infty, x]) = \mathcal{P}\left(X^{-1}(-\infty, x]\right), \text{ as } (-\infty, x] \in \mathcal{B}(\mathbb{R}) \ \forall \ x \in \mathbb{R}$$

$$\Rightarrow \mathcal{P}_X((-\infty, x]) = \mathcal{P}\left(\{\omega \in \Omega : X(\omega) \in (-\infty, x]\}\right) \quad [\text{Using } (4.31)]$$

$$\Rightarrow \mathcal{P}_X((-\infty, x]) = \mathcal{P}\left(\{\omega \in \Omega : X(\omega) \le x\}\right), \ x \in \mathbb{R} \ [\because y \in (-\infty, x] \Rightarrow y \le x] \ldots (4.33)$$

In continuous world, we have to obtain probability of the random variable X lying in different intervals very frequently. So, we use the following simple notation for (4.33).

$$\Rightarrow \mathcal{P}_X((-\infty, x]) = \mathcal{P}(X \le x), \ x \in \mathbb{R} \quad \ldots (4.34)$$

Now, we can define cumulative distribution function (CDF) or simply distribution function of a random variable as follows.

Cumulative Distribution Function (CDF): Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_X)$ be Borel measurable space. Then a function

$F_{\mathcal{P}_X} : \mathbb{R} \to [0, 1]$ defined by

$$\left.\begin{aligned} &F_{\mathcal{P}_X}(x) = \mathcal{P}_X\left((-\infty, x]\right), \ x \in \mathbb{R} \\ &\text{where } \mathcal{P}_X\left((-\infty, x]\right) = \mathcal{P}\left(X^{-1}(-\infty, x]\right), \ (-\infty, x] \in \mathcal{B}\left(\mathbb{R}\right) \ \forall \ x \in \mathbb{R} \end{aligned}\right\} \quad \ldots (4.35)$$

is called cumulative distribution function (CDF) or simply distribution function corresponding to the probability measure $\mathcal{P}_X$.

But by (4.34) we know that $\mathcal{P}_X\left((-\infty, x]\right) = \mathcal{P}\left(X \le x\right), \ x \in \mathbb{R}.$ So, (4.35) can be written as $F_{\mathcal{P}_X}(x) = \mathcal{P}_X\left((-\infty, x]\right) = \mathcal{P}\left(X \le x\right), \ x \in \mathbb{R}.$ … (4.36)

Note that (4.36) gives sum of all the probabilities up to and including x of the random variable X. Recall that in descriptive statistics sum of all frequencies up to a value of the variable are known as cumulative frequencies. Here instead of frequencies we are adding all the probabilities up to and including a particular value x of the random variable X so the world cumulative is added in front of distribution function. Cumulative distribution function (CDF) has a significance role in the all coming units of this course. You will realise it as we will proceed further in the course. Because of so frequent use of distribution function in probability theory, we will simply denote it by F_X instead of $F_{\mathcal{P}_X}$. So, using this notation (4.36) can be written as

$$F_X(x) = \mathcal{P}_X\left((-\infty, x]\right) = \mathcal{P}\left(X \le x\right), \ x \in \mathbb{R}. \quad \ldots (4.36a)$$

In view of (4.33) it can also be written as

$$F_X(x) = \mathcal{P}_X\left((-\infty, x]\right) = \mathcal{P}\left(\left\{\omega \in \Omega : X(\omega) \le x\right\}\right), \ x \in \mathbb{R}. \quad \ldots (4.36b)$$

Combining (4.35), (4.36), (4.36a) and (4.36b), we have

$$\left.\begin{aligned} F_{\mathcal{P}_X}(x) = F_X(x) &= \mathcal{P}_X\left((-\infty, x]\right) \\ &= \mathcal{P}\left(X^{-1}(-\infty, x]\right) \\ &= \mathcal{P}\left(\left\{\omega \in \Omega : X(\omega) \le x\right\}\right) \\ &= \mathcal{P}\left(X \le x\right), \ x \in \mathbb{R} \end{aligned}\right\} \quad \ldots (4.36c)$$

In view of (4.36a), we have

$$\mathcal{P}\left(a < X \le b\right) = \mathcal{P}\left(X \le b\right) - \mathcal{P}\left(X \le a\right) = F_X(b) - F_X(a)$$

$$\Rightarrow \mathcal{P}\left(a < X \le b\right) = F_X(b) - F_X(a) \quad \ldots (4.36d)$$

Also, we say that F_X is the distribution function of the random variable X instead of saying that corresponding to the probability measure $\mathcal{P}_X$. Further, if we have to deal with distribution functions of two or more random variables X and Y (say) then we denote their distribution functions by F_X and F_Y respectively. One more notation alert if we are dealing with only one random variable X in a problem then if we want we can simply use F instead of F_X. Keep all these notation issues in your mind. … (4.37)

Using cumulative distribution function (CDF) or simply distribution function we can obtain probabilities of the random variable X lying in a particular range of our interest. But to do so you should have good understanding of four properties of distribution function. But to prove these properties we need continuity of the probability measure. So, before discussing proofs of the properties of the CDF first we have to discuss what we mean by continuity of a probability measure which is discussed as follows.

Increasing Sequence of Events or Non-Decreasing Sequence of Events: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. A sequence $E_1, E_2, E_3, \ldots$ of events is said to be an increasing sequence of events or non-decreasing sequence of events if its members satisfy the following condition $E_1 \subseteq E_2 \subseteq E_3 \subseteq \ldots$ If $\bigcup_{n=1}^{\infty} E_n = E,$ then we say that the sequence $\{E_n\}_{n=1}^{\infty}$ increases to E and denote it by writing $E_n \uparrow E.$ The members of this sequence are visualised in Fig. 4.4 (a). … (4.38)

For example, consider the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_X)$ then $E_n = (-n, n), n = 1, 2, 3, \ldots$ is an increasing sequence of events in $\mathbb{R}$ and $E_n \uparrow \mathbb{R}.$

Decreasing Sequence of Events or Non-Increasing Sequence of Events: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. A sequence $E_1, E_2, E_3, \ldots$ of events is said to be decreasing sequence of events or non-increasing sequence of events if its members satisfy the following condition $E_1 \supseteq E_2 \supseteq E_3 \supseteq \ldots$ If $\bigcap_{n=1}^{\infty} E_n = E,$ then we say that the sequence $\{E_n\}_{n=1}^{\infty}$ decreases to E and denote it by writing $E_n \downarrow E.$ The members of this sequence are visualised in Fig. 4.4 (b). … (4.39)

For example, consider the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_X)$ then

$$E_n = \left(-\infty, x + \frac{1}{n}\right], n = 1, 2, 3, \ldots$$

So,

$$E_1 = (-\infty, x+1], E_2 = (-\infty, x+0.5], E_3 = \left(-\infty, x+\frac{1}{3}\right], E_4 = (-\infty, x+0.25], \ldots \text{ is}$$

a decreasing sequence of events in $\mathbb{R}$ and $E_n \uparrow (-\infty, x].$

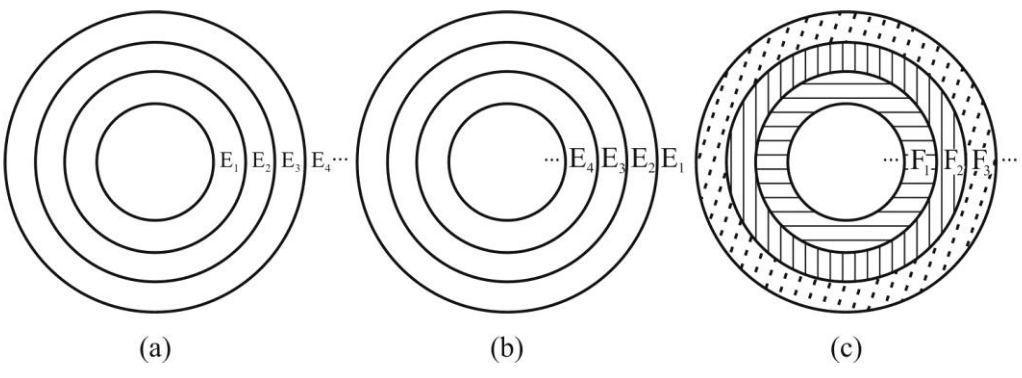

Fig. 4.4: Visualisation of the members of the (a) increasing sequence of events (b) decreasing sequence of events

From school mathematics you know that a function f is continuous at x = a if

$$\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x) = f(a), \ a \in \text{Domain of } f \qquad \ldots (4.40)$$

$$\text{or } LHL_{\text{at } x=a} = RHL_{\text{at } x=a} = \text{value of the function } f \text{ at } x = a. \qquad \ldots (4.41)$$

You may also refer (2.27) and (2.28) of the course MST-011.

You know that a probability measure $\mathcal{P}$ is a set function (refer 3.22 and 3.23 of the course MST-011 and 2.28 of this course), so members of the domain of $\mathcal{P}$ are events and its range is [0, 1], instead of the function $f : \mathbb{R} \to \mathbb{R}$, where both domain and codomain are real numbers. As usual like the function $f : \mathbb{R} \to \mathbb{R}$, we talk about continuity of the probability measure in terms of the members of its domain. But members of the domain of $\mathcal{P}$ are events so we will talk about the continuity of $\mathcal{P}$ in terms of events. Following two results discuss two types of continuity of a probability measure.

Continuity from Below of a Probability Measure: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. Suppose we have a sequence $E_1, E_2, E_3, \ldots$ of events with $E_n \uparrow E$, then $\lim_{n \to \infty} \mathcal{P}(E_n) = \mathcal{P}(E)$ or $\mathcal{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \lim_{n \to \infty} \mathcal{P}(E_n)$ $\left[\because \bigcup_{i=1}^{\infty} E_i = E\right]$.

That is probability measure respect limit from below. $\qquad \ldots (4.42)$

Proof: Since $E_n \uparrow E$, which means $E_1 \subseteq E_2 \subseteq E_3 \subseteq \ldots$ and

$$\bigcup_{n=1}^{\infty} E_n = E \qquad \ldots (4.43)$$

Let us define $F_1 = E_1, \ F_2 = E_2 \setminus E_1, F_3 = E_3 \setminus E_2, \ldots$ then we have (if you want you may refer Fig. 4.4 (c))

$$F_i \cap F_j = \phi, \text{ for } i \neq j \qquad \ldots (4.44)$$

$$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} F_n \qquad \ldots (4.45)$$

$$\bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{n} F_i \qquad \ldots (4.46)$$

Now, $\mathcal{P}(E) = \mathcal{P}\left(\bigcup_{n=1}^{\infty} E_n\right)$ [Using (4.43)]

$$\Rightarrow \mathcal{P}(E) = \mathcal{P}\left(\bigcup_{n=1}^{\infty} F_n\right) \qquad [\text{Using (4.45)}]$$

$$= \sum_{n=1}^{\infty} \mathcal{P}(F_n) \qquad \left[\begin{array}{l}\text{Using (4.44) and countable} \\ \text{additivity of probability measure}\end{array}\right]$$

$$= \lim_{n \to \infty} \sum_{i=1}^{n} \mathcal{P}(F_i) \qquad \left[\begin{array}{l}\text{Writing an infinite series of non-negative} \\ \text{terms in terms of its partial sum refer (4.63)} \\ \text{of the course MST-011}\end{array}\right]$$

$$= \lim_{n \to \infty} \mathcal{P}\left(\bigcup_{i=1}^{n} F_i\right) \left[\begin{array}{l}\text{Using finite additivity of probability measure} \\ \text{if you want you may refer (2.32) of this course}\end{array}\right]$$

$$= \lim_{n\to\infty} \mathcal{P}\left(\bigcup_{i=1}^{n} E_i\right) \quad [\text{Using } (4.46)]$$

$$= \lim_{n\to\infty} \mathcal{P}(E_n) \quad \left[\because E_1 \subseteq E_2 \subseteq E_3 \subseteq \ldots, \text{ so } \bigcup_{i=1}^{n} E_i = E_n\right]$$

This completes the proof. Second last line proves or part. But the main result is $\lim_{n\to\infty} \mathcal{P}(E_n) = \mathcal{P}(E)$. … (4.47)

Before discussing the continuity from above of a probability measure, first we have to state one result.

Result 1: $E_n \uparrow E$ if and only if $E_n^c \downarrow E^c$. … (4.48)

Proof: Proof of this result is beyond the scope of this course.

Continuity from Above of a Probability Measure: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. Suppose we have a sequence $E_1, E_2, E_3, \ldots$ of events with $E_n \downarrow E$, then $\lim_{n\to\infty} \mathcal{P}(E_n) = \mathcal{P}(E)$ or $\mathcal{P}\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mathcal{P}(E_n) \left[\because \bigcap_{n=1}^{\infty} E_n = E\right]$.

That is probability measure respect limit from above. … (4.49)

Proof: Since $E_n \downarrow E$, which means $E_1 \supseteq E_2 \supseteq E_3 \supseteq \ldots$ and

$$\bigcap_{n=1}^{\infty} E_n = E \qquad \ldots (4.50)$$

Using (4.42), we know that if $E_n \uparrow E$, then $\lim_{n\to\infty} \mathcal{P}(E_n) = \mathcal{P}(E)$. … (4.51)

Here we are given that $E_n \downarrow E$, so using (4.48) and (4.51), we have

$$\lim_{n\to\infty} \mathcal{P}(E_n^c) = \mathcal{P}(E^c)$$

$$\Rightarrow \lim_{n\to\infty} (1 - \mathcal{P}(E_n)) = 1 - \mathcal{P}(E) \quad \left[\because \text{for any event A}, \mathcal{P}(A^c) = 1 - \mathcal{P}(A)\right]$$

$$\Rightarrow \lim_{n\to\infty} 1 - \lim_{n\to\infty} \mathcal{P}(E_n) = 1 - \mathcal{P}(E)$$

$$\Rightarrow 1 - \lim_{n\to\infty} \mathcal{P}(E_n) = 1 - \mathcal{P}(E)$$

$$\Rightarrow - \lim_{n\to\infty} \mathcal{P}(E_n) = -\mathcal{P}(E)$$

$$\Rightarrow \lim_{n\to\infty} \mathcal{P}(E_n) = \mathcal{P}(E)$$

This completes the proof.

Remark 1: This result can also be proved following similar steps as in the proof of the result given by (4.42). You have to just interchange the role of union with intersection and accordingly doing some appropriate modification.

Now, we can state and prove four properties of CDF as follows.

Properties of CDF: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and F_X be the CDF of the random variable X then

(a) $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0.$ … (4.52)

(b) $F_X(\infty) = \lim_{x\to\infty} F_X(x) = 1.$ … (4.53)

(c) F_X is increasing or non-decreasing. … (4.54)

(d) F_X is right continuous. … (4.55)

Proof: By definition of CDF refer (4.36a), we have

$$F_X(x) = \mathcal{P}_X((-\infty, x]),\ x \in \mathbb{R} \quad \text{… (4.56)}$$

(a) $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = \lim_{x\to-\infty} \mathcal{P}_X((-\infty, x])$ [Using (4.56)]

$= \mathcal{P}_X\left(\bigcap_{x\to-\infty} ((-\infty, x])\right)$ $\left[\begin{array}{l}\because \mathcal{P}_X \text{ is a probability measure refer (4.27).} \\ \text{Also as } x\to-\infty, (-\infty, x] \text{ is a decreasing} \\ \text{sequence of events so, using (4.49)}\end{array}\right]$

$= \mathcal{P}_X(\phi)$

$= 0$ $[\because \mathcal{P}_X$ is a probability measure so $\mathcal{P}_X(\phi) = 0]$

(b) $F_X(\infty) = \lim_{x\to\infty} F_X(x) = \lim_{x\to\infty} \mathcal{P}_X((-\infty, x])$ [Using (4.56)]

$= \mathcal{P}_X\left(\bigcup_{x\to\infty} ((-\infty, x])\right)$ $\left[\begin{array}{l}\because \mathcal{P}_X \text{ is a probability measure refer (4.27).} \\ \text{Also as } x\to\infty, (-\infty, x] \text{ is an increasing} \\ \text{sequence of events so, using (4.42).}\end{array}\right]$

$= \mathcal{P}_X(\mathbb{R})$

$= 1$ $[\because \mathcal{P}_X$ is a probability measure so $\mathcal{P}_X(\mathbb{R}) = 1]$

(c) Let $x, y \in \mathbb{R}$ be such that $x \le y$. To prove that F_X is increasing or non-decreasing, we have to prove that $F_X(x) \le F_X(y)$.

Let $E = \{\omega \in \Omega : X(\omega) \le x\}$ and $F = \{\omega \in \Omega : X(\omega) \le y\}$. Since $x \le y \Rightarrow E \subseteq F$.

We know that every probability measure is monotonic refer (2.33) in Unit 2 of this course, therefore

$E \subseteq F$

$\Rightarrow \mathcal{P}(E) \le \mathcal{P}(F)$

$\Rightarrow \mathcal{P}(\{\omega \in \Omega : X(\omega) \le x\}) \le \mathcal{P}(\{\omega \in \Omega : X(\omega) \le y\})$

$\Rightarrow \mathcal{P}_X((-\infty, x]) \le \mathcal{P}_X((-\infty, y])$ [Using (4.33)]

$\Rightarrow F_X(x) \le F_X(y)$ [Using (4.56)]

Hence, F_X is increasing or non-decreasing function.

(d) To prove that F_X is right continuous at each $x \in \mathbb{R}$, we have to prove that $\lim_{h\to0^+} F_X(x+h) = F_X(x)$ or $\lim_{h\downarrow0} F_X(x+h) = F_X(x)$.

To prove it let ε_n be a decreasing sequence of positive terms such that $\varepsilon_n \to 0$ as $n \to \infty$. So, it is enough to prove that $\lim_{n\to\infty} F_X(x+\varepsilon_n) = F_X(x)$.

$\text{LHS} = \lim_{n\to\infty} F_X(x+\varepsilon_n) = \lim_{n\to\infty} \mathcal{P}(X \le x+\varepsilon_n)$ [Using (4.36c)]

$= \lim_{n\to\infty} \mathcal{P}(\{\omega \in \Omega : X(\omega) \le x+\varepsilon_n\})$ [Using (4.36c)]

$$= \mathcal{P}\left(\bigcap_{i=1}^{\infty}\{\omega \in \Omega : X(\omega) \le x + \varepsilon_i\}\right) \qquad \left[\begin{array}{l}\because \varepsilon_n \text{ is a decreasing sequence} \\ \text{and so using (4.49)}\end{array}\right]$$

$$= \mathcal{P}\left(\{\omega \in \Omega : X(\omega) \le x\}\right) \qquad [\because \varepsilon_n \to 0 \text{ as } n \to \infty]$$

$$= \mathcal{P}_X\left((-\infty, x]\right) \qquad [\text{Using (4.33)}]$$

$$= F_X(x) \qquad [\text{Using (4.56)}]$$

$$= \text{RHS}$$

Hence, F_X is right continuous at each $x \in \mathbb{R}$.

Now, you can try the following Self-Assessment Question.

SAQ 1

Check whether the function given by $F_X(x) = \begin{cases} 0, & x < 0 \\ 2 - e^{-x}, & x \ge 0 \end{cases}$ is a distribution function or not.

4.6 DISCRETE RANDOM VARIABLE

In Sec. 1.2 of Unit 1 of the course MST-011, you have gone through the definitions of discrete and continuous variables from a layman point of view. In Sec. 2.3 of Unit 2 of the same course MST-011, you have studied definition of a countable set which is a mathematical sound concept to understand the distinction between discrete and continuous variables. The purpose of introducing those concepts there was to give you a good exposure of countable sets so that you can understand the distinction between a discrete and a continuous random variable which we will discuss in this unit and continue throughout the course. Before studying this section and the next section of this unit it is recommended that first you should go through Sec. 1.2 and Sec. 2.3 of Units 1 and 2 respectively of the course MST-011. I am assuming that you have understood the meaning of countable set discussed in Sec. 2.3 because without understanding the meaning of countable set you cannot fully enjoy the definition of discrete random variable in true sense which is given as follows.

Definition of Discrete Random Variable: A random variable X is said to be discrete random variable if there exist a finite number of values $x_1, x_2, x_3, \ldots, x_n$ or countably infinite number of values $x_1, x_2, x_3, \ldots$ such that if $\Omega = \{x_1, x_2, x_3, \ldots, x_n\}$ or $\Omega = \{x_1, x_2, x_3, \ldots\}$ then $\mathcal{P}(\Omega) = 1$. … (4.57)

Also, the set of those $x_i \in \Omega$ which satisfy $\mathcal{P}(X = x_i) \ne 0$ is called the **support of the random variable** X. … (4.58)

Before considering some examples of discrete random variable let us also define probability mass function of a discrete random variable as follows.

Probability Mass Function (PMF) of a Discrete Random Variable: Let X be a discrete random variable. Probability mass function (PMF) of X is a function which associates unique probability to each value of X and 0 probability to the values which are not in the support of X. It is denoted by $p_X(x)$ where

$$p_X(x) = \begin{cases} \mathcal{P}(X = x), & \text{if } x \text{ is in support of } X \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (4.59)$$

$$\text{Note that } p_X(x) = \mathcal{P}(X = x) = \mathcal{P}(\{\omega \in \Omega : X(\omega) = x\}) \\ = \mathcal{P}(E), \qquad \ldots (4.60)$$

where $E = \{\omega \in \Omega : X(\omega) = x\}$ is an event so PMF actually associates probabilities to events which are expressed in terms of the values of the random variable X. But you know that the job of assigning probabilities to the events is done by the probability measure in a probability space. So, keeping this in view, we can also define PMF of a discrete random variable as follows.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be the probability space, then the probability measure $\mathcal{P}$ itself plays the role of PMF. … (4.61)

For a function $p_X(x)$ to be a **valid probability mass function** it has to satisfy the following two conditions.

$$\left.\begin{array}{ll} (1) & \textbf{Non-negativity}: \ p_X(x) \geq 0 \ \ \forall \ x \in \text{support of } X \\ (2) & \textbf{Normality}: \ \sum_{x \in S} p_X(x) = 1, \ \ \text{where S denotes support of } X \end{array}\right\} \qquad \ldots (4.62)$$

In the previous section, we have discussed cumulative distribution function (CDF) of a random variable in general and in detail. We have also discussed four properties of CDF. Here, we define CDF specifically for the discrete random variable as follows.

Cumulative Distribution Function of a Discrete Random Variable: Let X be a discrete random variable then CDF of X is denoted by F_X and is defined by

$$F_X(x) = \mathcal{P}(X \leq x). \qquad \ldots (4.63)$$

i.e., CDF in terms of PMF can be written as $F_X(x) = \sum_{x_k \leq x} \mathcal{P}(X = x_k)$ … (4.64)

Now, we discuss some examples to explain the idea of discrete random variable and its PMF and CDF.

Example 3: A tetrahedral die is thrown twice. Write sample space of this random experiment. If X denotes the maximum of the two outcomes, then

(a) obtain probability mass function of X.

(b) obtain CDF of X.

(c) make plot of PMF and CDF of X obtained in (a) and (b) to visualise them.

Solution: When a tetrahedral die is thrown twice then sample space is given by

$$\Omega = \{1, 2, 3, 4\}^2 = \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (2, 4), \\ (3, 1), (3, 2), (3, 3), (3, 4), (4, 1), (4, 2), (4, 3), (4, 4)\}.$$

Here, X denotes the maximum of the two outcomes. So, X can take values 1, 2, 3 and 4. Note that all 16 possible outcomes of this random experiment are equally likely, so appropriate probability measure on Ω is uniform probability measure. So, $\mathcal{P}(\{\omega_i\}) = \frac{1}{16}\ \forall\ \omega_i \in \Omega,\ i = 1, 2, 3, \ldots, 16.$ Now,

$$\mathcal{P}(X=1) = \mathcal{P}(\{\omega \in \Omega : X(\omega) = 1\}) = \mathcal{P}(\{(1, 1)\}) = \frac{1}{16}$$

$$\mathcal{P}(X=2) = \mathcal{P}(\{\omega \in \Omega : X(\omega) = 2\}) = \mathcal{P}(\{(1, 2), (2, 2), (2, 1)\}) = \frac{3}{16}$$

Similarly, $\mathcal{P}(X=3) = \mathcal{P}(\{(1, 3), (2, 3), (3, 3), (3, 1), (3, 2)\}) = \frac{5}{16}$ and

$$\mathcal{P}(X=4) = \mathcal{P}(\{(1, 4), (2, 4), (3, 4), (4, 4), (4, 1), (4, 2), (4, 3)\}) = \frac{7}{16}$$

(a) Probability mass function of X is given by

$$p_X(x) = \begin{cases} \frac{2x-1}{16}, & x = 1, 2, 3, 4 \\ 0, & \text{otherwise} \end{cases} \quad \left[\begin{array}{l} \because 1, 3, 5 \text{ and } 7 \text{ is an AP with first term } a = 1 \\ \text{and common difference } d = 2. \text{ So, } n^{th} \text{ term} \\ \text{of this AP is given by } a + (n-1)d = 2n - 1 \end{array}\right]$$

But all the time probabilities do not satisfy such a nice function. In those cases, we can write PMF like a piecewise function shown as follows.

$$p_X(x) = \begin{cases} 1/16, & \text{if } x = 1 \\ 3/16, & \text{if } x = 2 \\ 5/16, & \text{if } x = 3 \\ 7/16, & \text{if } x = 4 \\ 0, & \text{otherwise} \end{cases}$$

Sometimes probability mass function can also be described using tabular form and this form is known as probability distribution of the random variable X shown as follows.

X	1	2	3	4
$p_X(x)$	1/16	3/16	5/16	7/16

(b) By definition CDF of X is given by

$F_X(x) = \mathcal{P}(X \le x).$

$\therefore\ F_X(1) = \mathcal{P}(X \le 1) = \mathcal{P}(X = 1) = 1/16$

$F_X(2) = \mathcal{P}(X \le 2) = \mathcal{P}(X = 1) + \mathcal{P}(X = 2) = 1/16 + 3/16 = 4/16 = 1/4$

$$F_X(3) = \mathcal{P}(X \le 3) = \sum_{x=1}^{3} \mathcal{P}(X = x) = 1/16 + 3/16 + 5/16 = 9/16$$

$$F_X(4) = \mathcal{P}(X \le 4) = \sum_{x=1}^{4} \mathcal{P}(X = x) = 1/16 + 3/16 + 5/16 + 7/16 = 16/16 = 1$$

Generally, CDF is written like a piecewise function covering entire real line shown as follows.

$$F_X(x) = \begin{cases} 0, & \text{if } x < 1 \\ 1/16, & \text{if } 1 \le x < 2 \\ 4/16, & \text{if } 2 \le x < 3 \\ 9/16, & \text{if } 3 \le x < 4 \\ 1, & \text{if } x \ge 4 \end{cases}$$

(c) Graph of PMF is shown in Fig. 4.5 (a) and graph of CDF is shown in Fig. 4.5 (b).

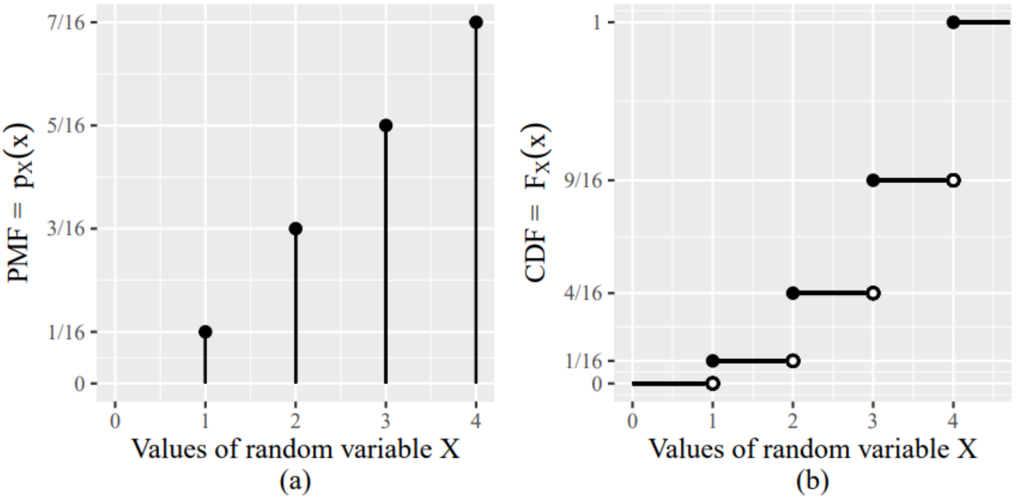

Fig. 4.5: Visualisation of (a) probability mass function (PMF) and (b) CDF, of the random variable X defined in Example 3

Example 4: Verify that CDF obtained in Example 3 satisfies all four conditions of a CDF.

Solution: CDF obtained in Example 3 is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < 1 \\ 1/16, & \text{if } 1 \le x < 2 \\ 4/16, & \text{if } 2 \le x < 3 \\ 9/16, & \text{if } 3 \le x < 4 \\ 1, & \text{if } x \ge 4 \end{cases}$$

We have to check that this CDF satisfies all the following four conditions mentioned as follows.

(a) $F_X(-\infty) = \lim\limits_{x \to -\infty} F_X(x) = 0.$

(b) $F_X(\infty) = \lim\limits_{x \to \infty} F_X(x) = 1.$

(c) F_X is increasing or non-decreasing.

(d) F_X is right continuous.

Let us check these conditions one at a time.

(a) $F_X(-\infty) = \lim\limits_{x \to -\infty} F_X(x) = 0.$ $\quad [\because x \to -\infty \Rightarrow x < 1 \text{ and for } x < 1, F(x) = 0]$

(b) $F_X(\infty) = \lim\limits_{x \to \infty} F_X(x) = 1.$ $\quad [\because x \to \infty \Rightarrow x > 4 \text{ and for } x \ge 4, F(x) = 1]$

(c) Let $x, y \in \mathbb{R}$ be such that $x \le y$. From the graph of the function F_X shown in Fig. 4.5 (b), it is obvious that $F_X(x) \le F_X(y)$. Still if you want to go in more detail then it can be done as follows.

If both x and y are < 1 then $F_X(x) = 0, F_X(y) = 0$, so $F_X(x) = F_X(y)$.

Similarly, if both x and y lie in the interval:

[1, 2) then $F_X(x) = F_X(y) = 1/16.$

[2, 3) then $F_X(x) = F_X(y) = 4/16.$

[3, 4) then $F_X(x) = F_X(y) = 9/16.$

$[4, \infty)$ then $F_X(x) = F_X(y) = 1.$

Finally, if x and y lie in different intervals among these interval under the condition that $x \le y,$ then $F_X(x) < F_X(y).$ Hence, $F_X(x) \le F_X(y)$ in all the possible cases. So, F_X is increasing or non-decreasing.

(d) Here F_X is a piece wise function and constant in each piece. We know that a constant function is continuous. So, to check that F_X is a right continuous function at each real number, we have to check its right continuity only at the points which separate two pieces. Keeping this in view, we have to check right continuity of F_X at the points x = 1, 2, 3 and 4 only.

Right continuity at $x = 1$

$\text{RHL}_{\text{at } x=1} = \lim\limits_{x \to 1^+} F_X(x) = \lim\limits_{x \to 1^+} (1/16) = 1/16,$ also $F_X(1) = 1/16.$

Right continuity at $x = 2$

$\text{RHL}_{\text{at } x=2} = \lim\limits_{x \to 2^+} F_X(x) = \lim\limits_{x \to 2^+} (4/16) = 4/16,$ also $F_X(2) = 4/16.$

Right continuity at $x = 3$

$\text{RHL}_{\text{at } x=3} = \lim\limits_{x \to 3^+} F_X(x) = \lim\limits_{x \to 3^+} (9/16) = 9/16,$ also $F_X(3) = 9/16.$

Right continuity at $x = 4$

$\text{RHL}_{\text{at } x=4} = \lim\limits_{x \to 4^+} F_X(x) = \lim\limits_{x \to 4^+} (1) = 1,$ also $F_X(4) = 1.$

Since right hand limit (RHL) is equal to value of the function at that point. So, F_X is a right continuous function.

Hence, F_X satisfies all the four conditions of CDF.

Remark 3: It did not happen by chance. Remember that it is a fact, all CDF's satisfy these conditions. We have already proved these properties of CDF, you may refer proofs of (4.52) to (4.55).

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

SAQ 2

Three babies born in a hospital in a day. Assume that there are only two possibilities boy or girl and both are equally likely. Ignore the possibility of transgender. If X denotes the number of girl babies born in the hospital on that day, then

(a) obtain PMF of X.

(b) obtain CDF of X.

(c) make plot of PMF and CDF of X obtained in (a) and (b) to visualise them.

SAQ 3

Explain the sizes of jumps at points 1, 2, 3 and 4 in the graph of CDF of X in

Fig, 4.5 (b) in Example 3. Is there any connection of these jumps at 1, 2, 3 and 4 with probabilities of the random variable at these points?

4.7 INDICATOR RANDOM VARIABLE

In Sec. 3.5 of Unit 3 of the course MST-011, you studied characteristic function refer (3.29) and (3.30). Recall that we denoted characteristic function of a set A by χ_A and also called indicator function of the set A. In the terminology of probability theory set A is replaced by an event E (say) and corresponding indicator function I_E is known as indicator random variable. A proper definition of indicator random variable is given as follows.

Definition of Indicator Random Variable: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and E be an event then indicator random variable $I_E : \Omega \to \{0, 1\}$ of event E is defined by

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

We will use the concept of indicator random variable and its properties in the further discussion of this course.

So, let us state and prove some special properties of indicator random variable as follows

Statement: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and E, F be any two events then prove that indicator random variables satisfy following properties.

(a) $(I_E)^n = I_E \quad \forall\ n \in \mathbb{N}$ … (4.65)

(b) $I_{E^c} = 1 - I_E$ … (4.66)

(c) $I_{E\cap F} = I_E I_F$ … (4.67)

(d) $I_{E\cup F} = I_E + I_F - I_E I_F$ … (4.68)

Proof: (a) Since indicator random variable assumes only two values 0 and 1 and clearly, $0^1 = 0, 0^2 = 0, 0^3 = 0, \ldots$ and $1^1 = 1, 1^2 = 1, 1^3 = 1, \ldots$ … (4.69)

So, we have $(I_E)^n = I_E \quad \forall\ n \in \mathbb{N}$

(b) We note that

$$\forall\ \omega \in \Omega, \text{ we have } 1 - I_E(\omega) = \begin{cases} 1-1, & \text{if } \omega \in E \\ 1-0, & \text{if } \omega \notin E \end{cases}$$

$$= \begin{cases} 0, & \text{if } \omega \in E \\ 1, & \text{if } \omega \notin E \end{cases}$$

$$= \begin{cases} 0, & \text{if } \omega \notin E^c \quad [\because \text{ if } \omega \in E \Rightarrow \omega \notin E^c] \\ 1, & \text{if } \omega \in E^c \quad [\because \text{ if } \omega \notin E \Rightarrow \omega \in E^c] \end{cases}$$

$$= I_{E^c}(\omega)$$

i.e., $1-I_E(\omega)=I_{E^c}(\omega) \quad \forall \ \omega \in \Omega$

$\Rightarrow 1-I_E=I_{E^c}$

(c) Let $\omega \in \Omega$, then four cases arise.

Case I: $\omega \in E$ and $\omega \in F$ but then $\omega \in E \cap F$

$$\therefore \ I_E(\omega)=1, \ I_F(\omega)=1 \text{ and } I_{E\cap F}(\omega)=1, \text{ so } I_E(\omega)\times I_E(\omega)=1\times 1=1=I_{E\cap F}(\omega) \quad \ldots (4.70)$$

Case II: $\omega \in E$ but $\omega \notin F \Rightarrow \omega \notin E \cap F$

$$\therefore \ I_E(\omega)=1, \ I_F(\omega)=0 \text{ and } I_{E\cap F}(\omega)=0, \text{ so } I_E(\omega)\times I_E(\omega)=1\times 0=0=I_{E\cap F}(\omega) \quad \ldots (4.71)$$

Case III: $\omega \notin E$ but $\omega \in F \Rightarrow \omega \notin E \cap F$

$$\therefore \ I_E(\omega)=0, \ I_F(\omega)=1 \text{ and } I_{E\cap F}(\omega)=0, \text{ so } I_E(\omega)\times I_E(\omega)=0\times 1=0=I_{E\cap F}(\omega) \quad \ldots (4.72)$$

Case IV: $\omega \notin E$ and $\omega \notin F \Rightarrow \omega \notin E \cap F$

$$\therefore \ I_E(\omega)=0, \ I_F(\omega)=0 \text{ and } I_{E\cap F}(\omega)=0, \text{ so } I_E(\omega)\times I_E(\omega)=0\times 0=0=I_{E\cap F}(\omega) \quad \ldots (4.73)$$

On combining (4.70) to (4.73), we have

$I_E(\omega)I_F(\omega)=I_{E\cap F}(\omega) \quad \forall \ \omega \in \Omega$

or $(I_E I_F)(\omega)=I_{E\cap F}(\omega) \ \forall \ \omega \in \Omega$ [Using (2.21) of the course MST-011]

$\Rightarrow I_E I_F = I_{E\cap F}$

(d)
$$\begin{aligned}
I_{E\cup F} &= 1-I_{(E\cup F)^c} && [\text{Using } (4.66)]\\
&= 1-I_{E^c\cap F^c} && [\text{Using Demorgon's law of set theory}]\\
&= 1-I_{E^c} I_{F^c} && [\text{Using } (4.67)]\\
&= 1-\left(1-I_E\right)\left(1-I_F\right) && [\text{Using } (4.66) \text{ again}]\\
&= 1-1+I_E+I_F-I_E I_F\\
&= I_E+I_F-I_E I_F
\end{aligned}$$

This completes the proof.

4.8 SUMMARY

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

- **Random Variable**: A random variable X is a function from a sample space (Ω) of a random experiment to a subset of the set of all real numbers, i.e., $X:\Omega \rightarrow \mathbb{R}$. So, the domain of a random variable is the sample space of a random experiment and the range may be any subset of the set of all real numbers.

- **Measurable and Borel Measurable Function**: Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces (refer 2.29 to recall what is a measurable space) then a function $f: \Omega_1 \to \Omega_2$ is said to be a **measurable function** with respect to $\mathcal{F}_1, \mathcal{F}_2$ if $f^{-1}(B) \in \mathcal{F}_1$ for each $B \in \mathcal{F}_2$. In particular, if $\Omega_2 = \mathbb{R},\ \mathcal{F}_2 = \mathcal{B}(\mathbb{R})$ then we say that f is **Borel measurable function**. .

- **Random Variable in Terminology of Measure Theory**: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be Borel measurable space. A function $X: \Omega \to \mathbb{R}$ is said to be a Borel measurable function with respect to $\mathcal{F}, \mathcal{B}(\mathbb{R})$ if $X^{-1}(B) \in \mathcal{F}$ for each $B \in \mathcal{B}(\mathbb{R})$.

- The **probability measure** $\mathcal{P}_X$ is known as **probability measure induced by the random variable X on the real line**. The probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}_X)$ with induced probability measure $\mathcal{P}_X$ is known as **induced probability space** by the random variable X on the real line, where $\mathcal{P}_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$ is defined by $\mathcal{P}_X(B) = \mathcal{P}(X^{-1}(B)),\ B \in \mathcal{B}(\mathbb{R})$

- **Cumulative Distribution Function**: CDF of the random variable X is a function $F_{\mathcal{P}_X} : \mathbb{R} \to [0, 1]$ and is defined by
$F_{\mathcal{P}_X}(x) = \mathcal{P}_X((-\infty, x]) = \mathcal{P}(X \le x),\ x \in \mathbb{R}.$.

- **Continuity from Below of a Probability Measure**: It states that
$$\lim_{n\to\infty} \mathcal{P}(E_n) = \mathcal{P}(E) \text{ or } \mathcal{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \lim_{n\to\infty} \mathcal{P}(E_n) \left[\because \bigcup_{i=1}^{\infty} E_i = E\right].$$

- **Continuity from Above of a Probability Measure**: It states that
$$\lim_{n\to\infty} \mathcal{P}(E_n) = \mathcal{P}(E) \text{ or } \mathcal{P}\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \mathcal{P}(E_n) \left[\because \bigcap_{n=1}^{\infty} E_n = E\right].$$

- **Properties of CDF**:

  (a) $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0.$

  (b) $F_X(\infty) = \lim_{x\to\infty} F_X(x) = 1.$

  (c) F_X is increasing or non-decreasing.

  (d) F_X is right continuous.

- **Discrete Random Variable**: A random variable X is said to be discrete random variable if there exist a finite number of values $x_1, x_2, x_3, \ldots, x_n$ or countably infinite number of values $x_1, x_2, x_3, \ldots$ such that if $\Omega = \{x_1, x_2, x_3, \ldots, x_n\}$ or $\Omega = \{x_1, x_2, x_3, \ldots\}$ then $\mathcal{P}(\Omega) = 1.$

- **Probability Mass Function of a Discrete Random Variable**: Probability mass function (PMF) of X is a function which associates unique probability to each value of X and 0 probability to the values which are not in the support of X. It is denoted by $p_X(x)$

- For a function $p_X(x)$ to be a **valid probability mass function** it has to satisfy the following two conditions.

(i) $p_X(x) \geq 0 \;\; \forall \; x \in \text{support of } X$

(ii) $\sum\limits_{x \in S} p_X(x) = 1$, where S denotes support of X..

- **Cumulative Distribution Function of a Discrete Random Variable**: Let X be a discrete random variable then CDF of X is denoted by F_X and is defined by $F_X(x) = \mathcal{P}(X \leq x)$.

4.9 TERMINAL QUESTIONS

1. If X is a discrete random variable, then does $\mathcal{P}(X \leq a) = \mathcal{P}(X < a) \;\; \forall \;\; a \in \mathbb{R}$?

2. If a variable X assumes value 'a' with probability 1. Can we consider it a random variable?

3. Using CDF of X obtained in part (b) of Example 3 find following probabilities.

4. With the help of an example show that more than one random variables can be defined on a sample space.

   (a) $\mathcal{P}(X \leq 2)$ (b) $\mathcal{P}(X < 2)$ (c) $\mathcal{P}(X = 2)$.

4.10 SOLUTIONS/ANSWERS

Self-Assessment Questions (SAQs)

1. We claim that given function is not cumulative distribution function of a random variable because
$\lim\limits_{x \to \infty} F_X(x) = \lim\limits_{x \to \infty}\left(2 - e^{-x}\right) = 2 - \lim\limits_{x \to \infty}\left(-e^{-x}\right) = 2 - 0 = 2 \neq 1.$

2. If b and g denote that new born baby is a boy and girl respectively, then sample space of three babies born in the hospital is given by

   $\Omega = \{bbb, bbg, bgb, gbb, bgg, gbg, ggb, ggg\}$.

   Here, X denotes the number of girl babies. So, X can attain values 0, 1, 2 and 3. Note that all the 8 possible outcomes of this random experiment are equally likely, so appropriate probability measure on Ω is uniform probability measure. So, $\mathcal{P}\left(\{\omega_i\}\right) = \frac{1}{8} \;\; \forall \;\; \omega_i \in \Omega, \;\; i = 1, 2, 3, \ldots, 8.$ Now,

$$\mathcal{P}(X = 0) = \mathcal{P}\left(\{\omega \in \Omega : X(\omega) = 0\}\right) = \mathcal{P}\left(\{bbb\}\right) = \mathcal{P}\left(\{b\} \cap \{b\} \cap \{b\}\right)$$

$$= \mathcal{P}\left(\{b\}\right)\mathcal{P}\left(\{b\}\right)\mathcal{P}\left(\{b\}\right) \begin{bmatrix} \because \text{If events E and F are independent} \\ \text{then } \mathcal{P}(E \cap F) = \mathcal{P}(E)\mathcal{P}(F) \end{bmatrix}$$

$$= \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

Similarly, $\mathcal{P}(X=1)=\mathcal{P}(\{\omega\in\Omega: X(\omega)=1\})=\mathcal{P}(\{bbg, bgb, gbb\})=\frac{3}{8}$

$\mathcal{P}(X=2)=\mathcal{P}(\{\omega\in\Omega: X(\omega)=2\})=\mathcal{P}(\{bgg, gbg, ggb\})=\frac{3}{8}$, and

$\mathcal{P}(X=3)=\mathcal{P}(\{\omega\in\Omega: X(\omega)=3\})=\mathcal{P}(\{ggg\})=\frac{1}{8}$

(a) Probability mass function of X is given by

$$p_X(x)=\begin{cases}1/8, & \text{if } x=0, 3\\ 3/8, & \text{if } x=1, 2\\ 0, & \text{otherwise}\end{cases}$$

(b) By definition, CDF of X is given by

$F_X(x)=\mathcal{P}(X\le x)$.

$\therefore\ F_X(0)=\mathcal{P}(X\le 0)=\mathcal{P}(X=0)=1/8$

$F_X(1)=\mathcal{P}(X\le 1)=\mathcal{P}(X=0)+\mathcal{P}(X=1)=1/8+3/8=4/8$

$F_X(2)=\mathcal{P}(X\le 2)=\sum_{x=0}^{2}\mathcal{P}(X=x)=1/8+3/8+3/8=7/8$

$F_X(3)=\mathcal{P}(X\le 3)=\sum_{x=0}^{3}\mathcal{P}(X=x)=1/8+3/8+3/8+1/8=8/8=1$

Generally, CDF is written like a piecewise function covering entire real line shown as follows.

$$F_X(x)=\begin{cases}0, & \text{if } x<0\\ 1/8, & \text{if } 0\le x<1\\ 4/8, & \text{if } 1\le x<2\\ 7/8, & \text{if } 2\le x<3\\ 1, & \text{if } x\ge 3\end{cases}$$

(c) Graph of PMF is shown in Fig. 4.6 (a) and graph of CDF is shown in Fig. 4.6 (b).

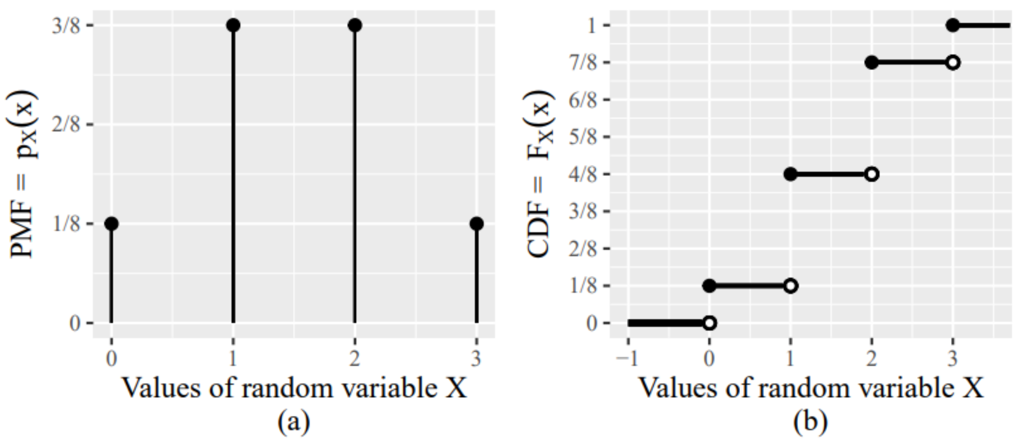

Fig. 4.6: Visualisation of (a) probability mass function (PMF) and (b) CDF, of the random variable X defined in SAQ 2

3. From Fig. 4.5 (a) and (b) we observe that there is a relationship between sizes of jumps at the values X = 1, 2, 3, 4 and probabilities at these values of X and this relationship is mentioned as follows and the same is visualised in Fig. 4.7.

Size of the jump at x = 1 in CDF of X is = $\mathcal{P}(X=1)=1/16.$

Similarly, size of the jump at x = 2 in CDF of X is = $\mathcal{P}(X=2)=3/16.$

Size of the jump at x = 3 in CDF of X is = $\mathcal{P}(X=3)=5/16.$

Size of the jump at x = 4 in CDF of X is = $\mathcal{P}(X=4)=7/16.$

This is also visualised in Fig. 4.7.

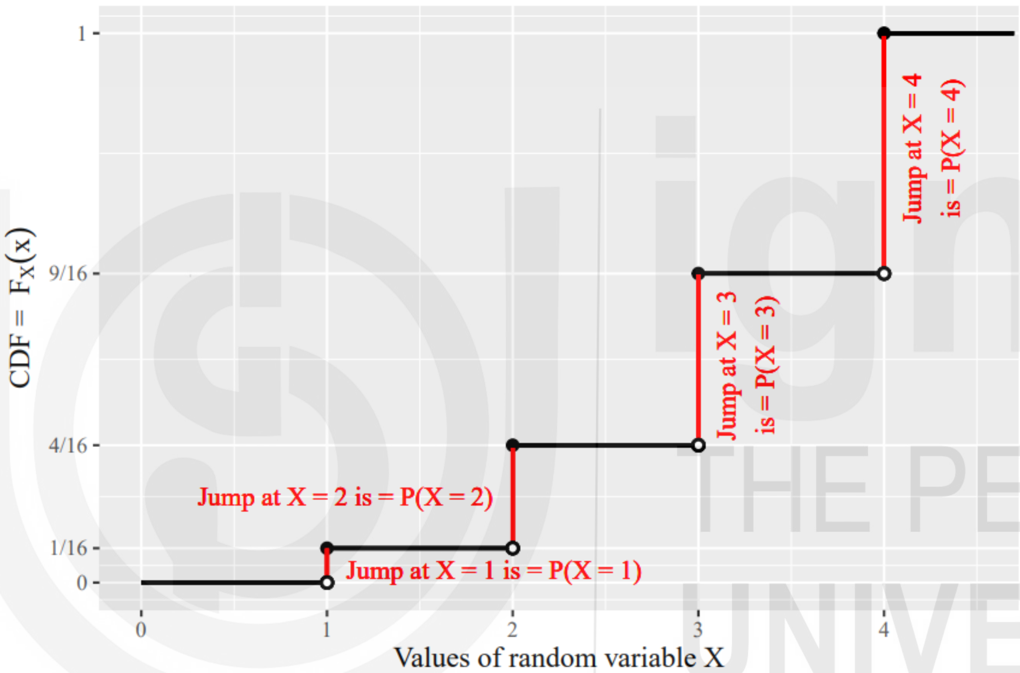

Fig. 4.7: Visualisation of the sizes of the jumps at the points X = 1, 2, 3 and 4 in the graph of CDF of the random variable X

Terminal Questions

1. The given statement $\mathcal{P}(X \le a) = \mathcal{P}(X < a)\ \forall\ a \in \mathbb{R}$ can hold only if $\mathcal{P}(X = a) = 0\ \forall\ a \in \mathbb{R}$. But in the case of discrete random variable, we do not have $\mathcal{P}(X = a) = 0\ \forall\ a \in \mathbb{R}$. Hence, in general $\mathcal{P}(X \le a) = \mathcal{P}(X < a)\ \forall\ a \in \mathbb{R}$ does not hold in the case of discrete random variable.

2. In this problem we are given that

$$\mathcal{P}(X = x) = \begin{cases} 1, & \text{if } x = a \\ 0, & \text{otherwise} \end{cases} \quad \ldots (4.74)$$

Here, the variable X is attaining only one value, so it seems that uncertainty factor is not involved as far as values that the variable X can take is concerned. But still, we treat it as a random variable because it satisfies definition of the random variable which simply states that

(i) X should assume real values only. This condition holds because a is a real number.

(ii) All probabilities should be non-negative, i.e., $\mathcal{P}(X = x) \geq 0 \ \forall \ x \in \mathbb{R}$ This condition also holds because of (4.74).

(iii) Sum of all probabilities should be 1. That is $\sum_{x} \mathcal{P}(X = x) = 1$. This condition also holds because $\mathcal{P}(X = a) + \mathcal{P}(X \neq a) = 1 + 0 = 1.$

3. Using (4.63), we have

$$\mathcal{P}(X \leq 2) = F_X(2) = \frac{4}{16}$$

$$\mathcal{P}(X < 2) = F_X(2^-) = \frac{1}{16} \left[\begin{array}{l} \because x < 2 \text{ means } x \neq 2 \text{ but } x \text{ can take value} \\ \text{very close to 2 which is denoted by } 2^-. \\ \text{So, } F_X(2^-) = \text{value of } F_X \text{ when } 1 \leq x < 2 \end{array}\right]$$

$$\mathcal{P}(X = 2) = \mathcal{P}(X \leq 2) - \mathcal{P}(X < 2) \left[\begin{array}{l} \text{For any a in the range of } X, \text{ we have} \\ \mathcal{P}(X \leq a) = \mathcal{P}(X < a) + \mathcal{P}(X = a) \\ \Rightarrow \mathcal{P}(X = a) = \mathcal{P}(X \leq a) - \mathcal{P}(X < a) \end{array}\right]$$

$$= \frac{4}{16} - \frac{1}{16} = \frac{3}{16} \qquad [\text{Using already obtained values}]$$

4. Let us consider the random experiment of tossing a coin twice. Sample space of this random experiment is $\Omega = \{HH, HT, TH, TT\}$. Let us define two random variables X and Y on Ω as follows.

X: denotes the number of heads, and

Y: denotes the number of tails before the first head.

So, we have

X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0 and

Y(HH) = 0, Y(HT) = 0, Y(TH) = 1, Y(TT) = 2.

Hence, probability distributions of random variable X is given by

X	0	1	2
$p_X(x)$	1/4	2/4	1/4

Hence, probability distributions of random variable X is given by

Y	0	1	2
$p_Y(y)$	2/4	1/4	1/4

Here, we have defined two random variables X and Y on the same sample space. In fact, we can also define more random variables on this sample space like Z denotes the number of tails, W denotes number of heads before the first tail, etc. Thus, we can say that more than one random variable can be defined on the same sample space. … (4.75)