UNIT 5

UNIVARIATE CONTINUOUS RANDOM VARIABLE

Structure

5.1 INTRODUCTION

In the previous unit, you studied what is a random variable and its visualisation as a function. You have also studied CDF which is an important concept in probability theory. After that, you studied PMF and CDF of a discrete random variable. In this unit first, we will define what is a continuous random variable in Sec. 5.2. Then we will discuss the probability density function (PDF) and CDF of a continuous random variable in Sec. 5.3.

What we have discussed in this unit is summarised in Sec. 5.4. 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. 5.5 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 Sec. 5.6.

In the next unit, you will study bivariate random variables.

Expected Learning Outcomes

After completing this unit, you should be able to:

- explain the concept of a continuous random variable;
- define probability density function, and CDF of a continuous random variable; and
- apply PDF and CDF to obtain probabilities of different intervals in the range of continuous random variable X.

5.2 DEFINITIONS OF CONTINUOUS RANDOM VARIABLE

From the discussion of Sec. 1.2 of Unit 1 and Sec. 2.3 of Unit 2 of the course MST-011 and the previous unit, you have understood that if a variable attains finite or countably infinite values then it falls in the category of discrete random variables. You also know that values of a discrete variable can be written in a sequence. But if it is not possible to write all possible values of a variable in a sequence then such a variable falls in the category of continuous variable. In other words, we can say that the list of possible values of a continuous variable is **uncountable** in nature. … (5.1)

So, all the random experiments where possible outcomes are not countable or uncountable are studied using continuous random variables. For example,

(i) The variable X which represents height of males in India who are 20 years old (here we have taken 20 years just because after 20 years generally height does not increase) or more is a continuous variable because there is no limit of precision. The height of a male may be 165.85674394257643789 cm, etc. So, answers to different questions like what is the probability that the height of a randomly selected male from this group lies in the interval 162 cm to 166 cm can be addressed by forming a distribution of the continuous random variable X. We will discuss some such problems in Unit 14 of this course.

(ii) Suppose you have a group of people who are so skilled that all their dart throws hit inside and on the unit circle $C1$ shown in Fig. 5.1 (a). Let X denote the distance of the point where dart hits inside or on the boundary of the circle to the centre of the circle $C1.$ If you are interested in the probability that $X < 1/2.$ That is X lies inside the circle $C2$ (refer Fig. 5.1 b) with radius 1/2 and having the same centre as that of $C1,$ then it will be a continuous random variable and can take any value between 0 and 1 including 0 and 1.

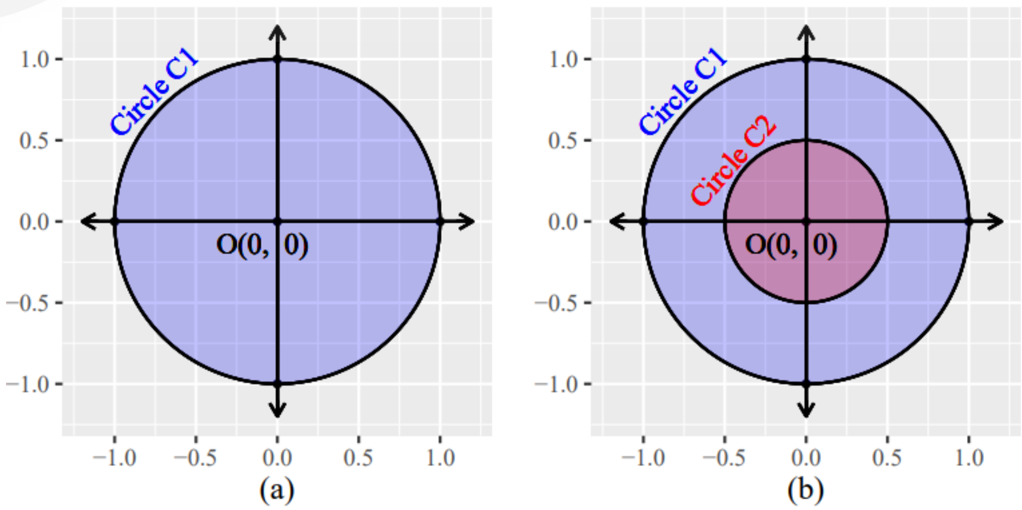

Fig. 5.1: Visualisation of (a) a unit circle (b) a circle of radius 1/2 inside the unit circle

So, here possible outcomes in both the experiments mentioned in (i) and (ii) are uncountable and therefore, to study such experiments, we should have an understanding of the distribution function of a continuous random variable.

With this introductory remark on the nature of a continuous random variable, let us define it. But here, we will discuss two definitions the first one will be very straightforward and the second will be from measure theory point of view. You should understand both definitions to get the real flavour of the concept.

First Definition of Continuous Random Variable: A random variable X falls in the category of continuous random variable if its CDF is differentiable almost everywhere. Almost everywhere means the measure of the set where it is not differentiable is zero. … (5.2)

Refer to Fig. 4.5 (b) of the previous unit where we have plotted the graph of CDF of a discrete random variable discussed in Example 3 of the same unit. Note that this graph has some jumps, in particular, at the points X = 1, 2, 3 and 4. So, it is not continuous at these points hence it is not differentiable at these points. Since we know that if a function is not continuous at a point, then it will not be differentiable at that point. But CDF of a continuous random variable is a smooth function without any jump refer to Fig. 5.4 (b) where we have shown CDF of the continuous random variable discussed in Example 2. … (5.3)

Second Definition of Continuous Random Variable: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and $(\Omega, \mathcal{B}(\mathbb{R}), \mathcal{P}_X)$ be induced probability space. A random variable X is said to be a continuous random variable if

$\mathcal{P}_X(B) = 0 \ \forall \ B \in \mathcal{B}(\mathbb{R})$ whenever $\lambda(B) = 0$, where λ is a Lebesgue measure.

i.e., the induced probability measure $\mathcal{P}_X$ assigns zero probability to all the members of the Borel σ-field $\mathcal{B}(\mathbb{R})$ which has zero Lebesgue measure. (5.4)

You know that Lebesgue measure of all types of intervals is their length refer to (3.72). So, singleton sets have zero Lebesgue measure because

Length of {x} = length of [x, x] = x – x = 0. … (5.5)

So, the probability that a continuous random variable X is equal to a particular value x (say) is zero. That is $\mathcal{P}(X = x) = 0 \ \ \forall \ \ x \in \mathbb{R}$. … (5.6)

We have also proved (5.6) earlier in general, if you want you may refer to (3.86). The result mentioned in (5.6) is special in the sense that it is true only for continuous random variables but not true for discrete random variables. … (5.7)

Example 1: In each of the following parts identify whether the given variable is discrete or continuous by nature.

(a) Number of claims an insurance company may receive during a month.

(b) Waiting time for a metro in Delhi after reaching on the platform.

(c) A variable assumes values $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots$

Solution: To identify the nature of these variables, we have to look at their possible values. If possible, values that a variable may attain can be written in a sequence then it will be discrete by nature, but, if possible, values that a variable may attain, cannot be written in a definite pattern (sequence) then the corresponding variable will be continuous by nature. Now, let us check it for each part.

(a) The number of claims that an insurance company may receive during a month may be 0, 1, 2, 3, 4, … These possible values form a sequence and hence the corresponding variable falls in the category of discrete random variables.

(b) Waiting time (in seconds, say) for a metro in Delhi after reaching the platform starts from 0 but after zero you cannot write the next possible value because it may be 0.1 or 0.01 or 0.001 or 0.0001 and so on. So, in this case, possible values of the random variable cannot be written in a sequence. Hence, the waiting time variable is continuous by nature. Keep in mind that in such cases, we think from a mathematical point of view not in terms of the time measurement instrument we have. For example, generally, we measure time in the nearest minutes or at the most seconds. If you will think this point of view then this variable will become discrete. But to judge the nature of the variable we think mathematically.

(c) Here values of the variable follow a definite pattern and following the same pattern we can easily write successive terms of the variable. For example, the next three values of the variable are $\frac{5}{6}, \frac{6}{7}, \frac{7}{8}.$ Hence, like part (a) it also falls in the category of discrete variables.

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

SAQ 1

Identify the nature of the variable “Exact weights of half litre milk packets” of a particular brand as discrete or continuous.

5.3 PDF AND CDF OF CONTINUOUS RANDOM VARIABLE

In the previous unit, you studied discrete random variable and its probability mass function (PMF) and cumulative distribution function (CDF). In this section, we will discuss such functions for a continuous random variable. In the continuous world, we have a probability density function (PDF) instead of PMF in the discrete world. However, like the discrete world, CDF has the same name in the continuous world as well. Let us first define CDF and then PDF as follows.

Cumulative Distribution Function of a Continuous Random Variable: Let X be a continuous random variable then CDF of X is denoted by F_X and is defined by $F_X(x) = \mathcal{P}(X \le x).$ That is like the discrete case CDF gives accumulated probability up to and including x. … (5.8)

Recall from Sec. 4.6 of the previous unit in the discrete case, we have used summation to accumulate all the probabilities up to and including x, you may refer to (4.64). In Sec. 5.2 of Unit 5 of the course MST-011, you have studied that the job of summation is done by integration in the continuous world and how it is done all this has been explained and visualised in Sec. 5.2 of the course MST-011. But before expressing CDF in terms of integration first, we have to define another function known as probability density function (PDF) which is defined as follows. … (5.9)

Probability Density Function of a Continuous Random Variable: Let X be a continuous random variable. The random variable X will have a probability density function (PDF) f_X if there exists a Borel measurable function (refer to 4.10 to recall what is a Borel measurable function) $f_X : \mathbb{R} \to [0, \infty]$ such that

$$\mathcal{P}_X(B) = \mathcal{P}(X \in B) = \int_B f_X(x)\,dx,\ B \in \mathcal{B}(\mathbb{R}) \quad \text{… (5.10)}$$

In particular, if B = (a, b], then

$$\mathcal{P}_X(B) = \mathcal{P}(X \in (a, b]) = \int_a^b f_X(x)\,dx,\ B = (a, b] \in \mathcal{B}(\mathbb{R}) \quad \text{… (5.11)}$$

To recall the meaning of the notation used in LHS of (5.11), you may refer (4.31) and (4.32) of the previous unit. … (5.12)

In Sub Sec. 5.2.2 of Unit 5 of the course MST-011, you have studied how definite integral gives area bounded by four curves: (i) two corresponding to lower and upper limits (ii) third is corresponding to the function which we are going to integrate, here it is f_X (iii) fourth is corresponding to the variable of integration. So, the area under the curve of the function f_X gives probability. Therefore, to be a **valid PDF** it must satisfy the following two conditions.

(a) $f_X(x) \geq 0 \quad \forall \quad x \in \mathbb{R} \left[\because \text{Probabilities are always non-negative}\right]$ … (5.13)

(b) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1 \quad \begin{bmatrix} \because \text{Sum of all probabilities should be 1 and in} \\ \text{continuous world sum is given by integration} \end{bmatrix}$ … (5.14)

In (4.64), we expressed CDF of a discrete random variable X in terms of its PMF. Using (4.36a), (5.8) and (5.11), we can express CDF of a continuous random variable X in terms of its PDF as follows.

$$F_X(x) = \mathcal{P}(X \leq x) = \mathcal{P}(X \in (-\infty, x]) = \int_{-\infty}^{x} f_X(x)\,dx \quad \text{... (5.15)}$$

$$= \int_{-\infty}^{x} f_X(t)\,dt \quad \text{... (5.16)}$$

$$\begin{bmatrix} \because \text{ within the integral sign x is a dummy variable,} \\ \text{if you like, you can use any other variable in place of x} \end{bmatrix}$$

In the discussion of this unit, Unit 6, partially in Unit 7 and Units 13 to 16 of this course, we will be dealing with PDF and CDF most of the time. So, it is right time to make good friendship with PDF and CDF. So, let us spend some time with our two new friends to know more about them. So far you have understood the kind of job they do in probability theory. Three more things to know about them will be sufficient for us to complete the journey of this course smoothly mentioned as follows.

(a) Connection between PDF and CDF to evaluate probability of an event

(b) Visualisation of the Connection between PDF and CDF in terms of their way of giving the probability of an event

(c) Interpretation of PDF

Let us discuss these taken one at a time.

(a) Connection between PDF and CDF to evaluate probability of an event

Recall that in discrete case, we were obtaining probabilities using PMF. Similarly, in continuous case using (5.11), we can obtain probabilities of an event of our interest using PDF. But as you will observe in solutions of Example 2 and SAQ 2, in continuous case, it is easy to evaluate required probability using CDF instead of PDF. Reason for this is: to obtain required probability using PDF we have to evaluate integral but if we do the same job using CDF then we have to evaluate value of a function (CDF) at two points. We know and also feel it when we will deal with solution of the examples and SAQs that generally it is easy to evaluate the values of a function at two points instead of evaluating the integral. So, in a continuous case to obtain probability of our interest, we will prefer to work with CDF instead of PDF. So, you should keep this important observation in mind whenever you go for working in a continuous case. Before discussing more about the role of PDF and CDF to evaluate the probability of an event, you should also keep the following important observations in your mind regarding the continuous case.… (5.17)

(1) $F_X(b) = \mathcal{P}(-\infty < X \le b) = \mathcal{P}((-\infty < X \le a] \cup (a, b]),\ \text{where } a < b$

$$= \mathcal{P}(-\infty < X \le a) + \mathcal{P}(a < X \le b) \quad [\text{Using } (2.32)]$$

$$\Rightarrow \mathcal{P}(a < X \le b) = \mathcal{P}(-\infty < X \le b) - \mathcal{P}(-\infty < X \le a)$$

$$= F_X(b) - F_X(a) \quad [\text{Using } (5.15)]$$

$$\Rightarrow \mathcal{P}(a < X \le b) = F_X(b) - F_X(a) \quad \ldots (5.18)$$

(2) $\mathcal{P}(a \le X \le b) = \mathcal{P}(X = a) + \mathcal{P}(a < X \le b)$

$$= 0 + \mathcal{P}(a < X \le b) \quad [\text{Using } (5.6)]$$

$$= \mathcal{P}(a < X \le b)$$

$$= F_X(b) - F_X(a) \quad [\text{Using } (5.18)]$$

$$\Rightarrow \mathcal{P}(a \le X \le b) = F_X(b) - F_X(a) \quad \ldots (5.19)$$

(3) $\mathcal{P}(a < X < b) = \mathcal{P}(a < X \le b) - \mathcal{P}(X = b)$

$$= \mathcal{P}(a < X \le b) - 0 \quad [\text{Using } (5.6)]$$

$$= \mathcal{P}(a < X \le b)$$

$$= F_X(b) - F_X(a) \quad [\text{Using } (5.18)]$$

$$\Rightarrow \mathcal{P}(a < X < b) = F_X(b) - F_X(a) \quad \ldots (5.20)$$

(4) $\mathcal{P}(a \le X < b) = \mathcal{P}(X = a) + \mathcal{P}(a < X < b)$

$$= \mathcal{P}(X = a) + \mathcal{P}(a < X \le b) - \mathcal{P}(X = b)$$

$$= 0 + \mathcal{P}(a < X \le b) - 0 \quad [\text{Using } (5.6)]$$

$$= \mathcal{P}(a < X \le b)$$

$$= F_X(b) - F_X(a) \quad [\text{Using } (5.18)]$$

$$\Rightarrow \mathcal{P}(a \le X < b) = F_X(b) - F_X(a) \quad \ldots (5.21)$$

From (5.11) and (5.18) to (5.21), we have

$$\mathcal{P}(a \le X < b) = \mathcal{P}(a \le X \le b) = \mathcal{P}(a < X < b) = \mathcal{P}(a < X \le b)$$
$$= \int_a^b f_X(x)\,dx = F_X(b) - F_X(a) \quad \dots (5.22)$$

Remember that this holds only in the case of continuous random variable. This completes the discussion of (a). … (5.23)

(b) Visualisation of the Connection between PDF and CDF in terms of their way of giving the probability of an event

To visualise the connection between PDF and CDF let us consider a normal distribution with mean 50 and standard deviation 2. From (5.18) of Unit 5 of the course MST-011 and in view of (5.22) of this unit, we can say that shaded region in blue colour in Fig. 5.2 (a) gives probability that random variable X lies between 48 and 50. So, area under PDF gives probability that random variable X lies in an interval and here interval is (48, 50) or (48, 50] or [48, 50) or [48, 50]. Also, the difference between the heights of CDF at 50 and 48 gives the same probability which was given by the shaded region in blue colour under PDF. That is

$$DC = EC - ED = EC - AB = F_X(50) - F_X(48) = \int_{48}^{50} f_X(x)\,dx \quad \dots (5.24)$$

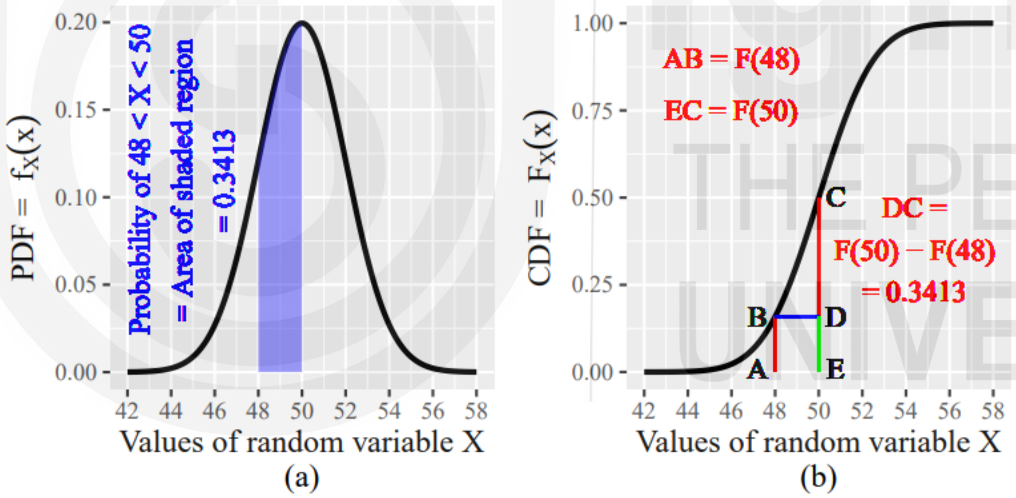

Fig. 5.2: Visualisation of probability (a) as area under PDF where 48 < X < 50 (b) as the difference of two values of CDF at X = 50 and X = 48

This connection between PDF and CDF can be seen in Fig. 5.2 (a) and (b). If the integral $\int_{48}^{50} f_X(x)\,dx$ and $F_X(50) - F_X(48)$ both give the probability $\mathcal{P}(48 \le X < 50)$ or $\mathcal{P}(48 \le X \le 50)$ or $\mathcal{P}(48 < X < 50)$ or $\mathcal{P}(48 < X \le 50)$ then we will prefer to use the second option $F_X(50) - F_X(48)$ because as we have already mentioned that generally to evaluate the value of a function at two points and then take their difference is easy compare to evaluate the integral over that interval. … (5.25)

(c) Interpretation of PDF

Like PMF $p_X(x)$ of a discrete random variable PDF $f_X(x)$ also satisfies the first condition that $f_X(x) \ge 0$ for all $x \in \mathbb{R}$. But outputs of $f_X(x)$ at different values of x are not probabilities whereas outputs of $p_X(x)$ were probabilities. In fact, outputs of $f_X(x)$ at some values of x may be even more than 1 which was not possible in the case of PMF. So, if outputs of $f_X(x)$ are not probabilities, then what they are? Let us proceed to answer this question as follows.

Consider a very tiny interval $\left[x-\frac{\varepsilon}{2}, x+\frac{\varepsilon}{2}\right]$ of length $\varepsilon > 0$ around a value x of X. Now, by definition of PDF probability that random variable X lies in the interval $\left[x-\frac{\varepsilon}{2}, x+\frac{\varepsilon}{2}\right]$ is given by integrating $f_X(x)$ over the interval $\left[x-\frac{\varepsilon}{2}, x+\frac{\varepsilon}{2}\right]$, so, we have

$$\mathcal{P}\left(\left[x-\frac{\varepsilon}{2}, x+\frac{\varepsilon}{2}\right]\right) = \int_{x-\frac{\varepsilon}{2}}^{x+\frac{\varepsilon}{2}} f_X(x)\,dx \left[\begin{array}{l}\text{Due to (5.22), it does not matter whether}\\ \text{we take} < \text{or} \le \text{sign in } x-\frac{\varepsilon}{2} < X < x+\frac{\varepsilon}{2}\end{array}\right] \quad \ldots (5.26)$$

Now, as $\varepsilon \to 0^+$ or $\varepsilon \downarrow 0$ then in the tiny interval $\left[x-\frac{\varepsilon}{2}, x+\frac{\varepsilon}{2}\right]$, PDF $f_X(x)$ will be almost constant. Let $f_X(x) = k,$ for $x \in \left[x-\frac{\varepsilon}{2}, x+\frac{\varepsilon}{2}\right].$ … (5.27)

So, (5.26) reduces to

$$\mathcal{P}\left(\left[x-\frac{\varepsilon}{2}, x+\frac{\varepsilon}{2}\right]\right) = k\left(x+\frac{\varepsilon}{2}-x+\frac{\varepsilon}{2}\right)\left[\because \int_a^b k\,dx = k(b-a)\right]$$

$$\Rightarrow \mathcal{P}\left(\left[x-\frac{\varepsilon}{2}, x+\frac{\varepsilon}{2}\right]\right) = k\,\varepsilon \quad \ldots (5.28a)$$

$$\Rightarrow k = \frac{\mathcal{P}\left(\left[x-\frac{\varepsilon}{2}, x+\frac{\varepsilon}{2}\right]\right)}{\varepsilon}$$

$$\Rightarrow \text{Value of PDF at a point} = \frac{\text{Probability over the tiny interval}}{\text{Length of the same tiny interval}\left[x-\frac{\varepsilon}{2}, x+\frac{\varepsilon}{2}\right]} \quad \ldots (5.28b)$$

So, as $\varepsilon \to 0^+$ or $\varepsilon \downarrow 0$ then LHS of (5.28b) can be approximately considered as value of the probability density function $f_X(x)$ at X = x and value of the ratio in RHS of (5.28b) can be considered as probability per unit length around the point X = x.

Hence, value of probability density function at a point X = x can be interpretated as **probability per unit length around the point X = x**.

In Unit 14, you will study normal distribution. Since so far you have not studied normal distribution so let us evaluate the value of LHS and RHS of (5.28b) using R around the point X = 49 and taking $\varepsilon = 0.002.$ Screenshot of R codes and their outputs is shown as follows. … (5.29)

```
> dnorm(x = 49,mean = 50,sd = 2) # LHS gives value of pdf at the point X = 49
[1] 0.1760327
> (pnorm(49+0.001,50,2)-pnorm(49-0.001,50,2))/0.002 # Value of RHS of (5.28)
[1] 0.1760327
```

So, density at the point X = 49 is 0.1760327. Let us obtain density at the point X = 54 also. Screenshot of R codes and their outputs is shown as follows.

```
> dnorm(x = 54, mean = 50, sd = 2) # gives value of pdf at the point X = 54
[1] 0.02699548
> (pnorm(54+0.001, 50, 2) - pnorm(54-0.001, 50, 2))/0.002 # Value of RHS of (5.28)
[1] 0.02699549
```

So, density at the point X = 54 is 0.0269954. Both densities at X = 49 and 54 are visualised in Fig. 5.3 (a). Note that density at X = 49 is more than at X = 54. In fact, density at X = 49 is 6.520819 times than that of at X = 54.

$$\left[\because \frac{\text{Density at } X=49}{\text{Density at } X=54} = \frac{0.1760327}{0.0269954} = 6.520819\right]$$

Higher density means probability that values of X lie near X = 49 is higher compare to the probability that values of X lie near X = 54 which is visualised in Fig. 5.3 (b) by considering probabilities of the intervals (48.5, 49.5) and (53.5, 54.5) each of length 1 unit. Screenshot of R code and outputs to get these probabilities is given as follows. … (5.30)

```
> pnorm(49.5, 50, 2) - pnorm(48.5, 50, 2)
[1] 0.1746663
> pnorm(54.5, 50, 2) - pnorm(53.5, 50, 2)
[1] 0.02783468
```

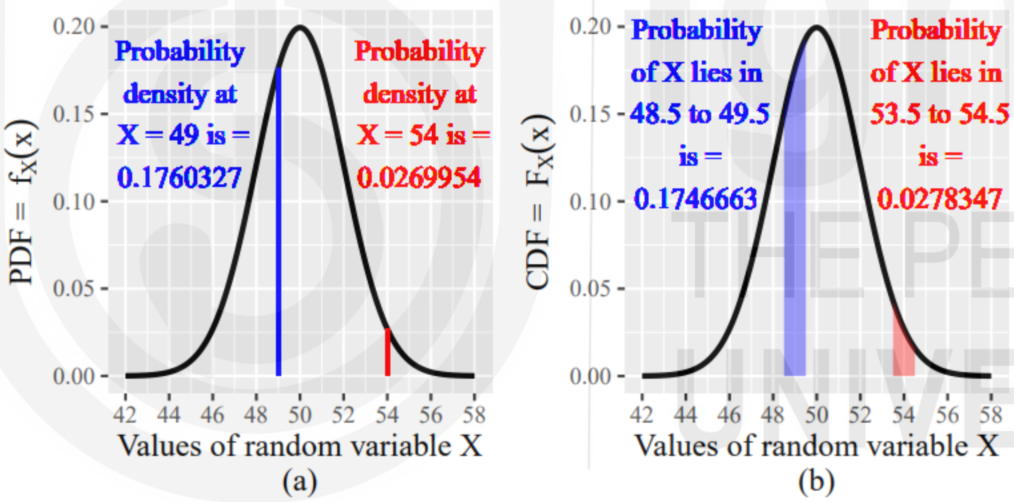

Fig. 5.3: Visualisation of (a) density at X = 49 and 54 (b) comparison of probabilities around X = 49 and 54

Now, we discuss some examples to explain the calculation of probabilities when PDF or CDF of continuous random variable is given to us.

Example 2: **For given PDF obtain CDF and probabilities**: Probability density function of a random variable X is given by

$$f_X(x) = \begin{cases} k\,x(1-x), & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases} \quad \ldots (5.31)$$

(a) Find value of k so that $f_X(x)$ becomes a valid PDF. Also, plot graph of PDF.

(b) Obtain CDF of X and plot its graph.

(c) Find probability that random variable X lies between (i) 1/4 and 1/2 (ii) 1/3 and 1 (iii) -3 and 2 (iv) 0.1 and 1.1

Solution: Since $x(1-x) > 0$, for $0 < x < 1$. So, to become a valid density, we must have k > 0.

(a) We can obtain value of k so that $f_X(x)$ becomes a valid PDF by solving

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1 \qquad \left[\text{Using } (5.14)\right]$$

$$\Rightarrow \int_{-\infty}^{0} f_X(x)\,dx + \int_{0}^{1} f_X(x)\,dx + \int_{1}^{\infty} f_X(x)\,dx = 1$$

$$\Rightarrow \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} k\,x(1-x)\,dx + \int_{1}^{\infty} 0\,dx = 1 \qquad \left[\text{Using } (5.31)\right]$$

$$\Rightarrow 0 + k\int_{0}^{1} x(1-x)\,dx + 0 = 1 \Rightarrow k\int_{0}^{1} x^{2-1}(1-x)^{2-1}\,dx = 1$$

$$\Rightarrow k\,B(2,\,2) = 1 \qquad \left[\text{Using } (8.8) \text{ of the course MST-011}\right]$$

$$\Rightarrow k\frac{\lceil 2 \lceil 2}{\lceil 4} = 1 \qquad \left[\text{Using } (8.21p5) \text{ of the course MST-011}\right]$$

$$\Rightarrow k\frac{\lfloor 1 \lfloor 1}{\lfloor 3} = 1 \qquad \left[\text{Using } (8.21p3) \text{ of the course MST-011}\right]$$

$$\Rightarrow k = 6 \qquad \ldots (5.32)$$

Graph of PDF is shown in Fig. 5.4 (a).

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

$$F_X(x) = \mathcal{P}\left(X \le x\right) = \int_{-\infty}^{x} f_X(x)\,dx = \int_{-\infty}^{x} 6\,x(1-x)\,dx \qquad \ldots (5.33)$$

Following three cases arise.

Case I: When x < 0. So, (5.33) implies

$$F_X(x) = \int_{-\infty}^{x} 0\,dx = 0 \qquad \ldots (5.34)$$

Case II: When $0 \le x < 1.$ So, (5.33) implies

$$F_X(x) = \int_{-\infty}^{x} 6\,x(1-x)\,dx = \int_{0}^{x} 6\,x(1-x)\,dx \quad \left[\because \text{ here } x \ge 0\right]$$

$$= 6\int_{0}^{x}\left(x - x^2\right)dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^x = 6\left(\frac{x^2}{2} - \frac{x^3}{3} - 0\right) = 3x^2 - 2x^3 \ldots (5.35)$$

Case III: When $x \ge 1.$ So, (5.33) implies

$$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} 6\,x(1-x)\,dx + \int_{1}^{x} 0\,dx \quad \left[\begin{array}{l}\because \text{ here } x \ge 1 \\ \text{and using } (5.31)\end{array}\right]$$

$$= \left[3x^2 - 2x^3\right]_0^1 \qquad \left[\text{Using } (5.35)\right]$$

$$= 3 - 2 - 0 = 1 \qquad \ldots (5.36)$$

Combining (5.34) to (5.36), we get

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 3x^2 - 2x^3, & \text{if } 0 \le x < 1 \\ 1, & \text{if } x \ge 1 \end{cases} \qquad \dots (5.37)$$

Finally, graph of CDF is shown in Fig. 5.4 (b).

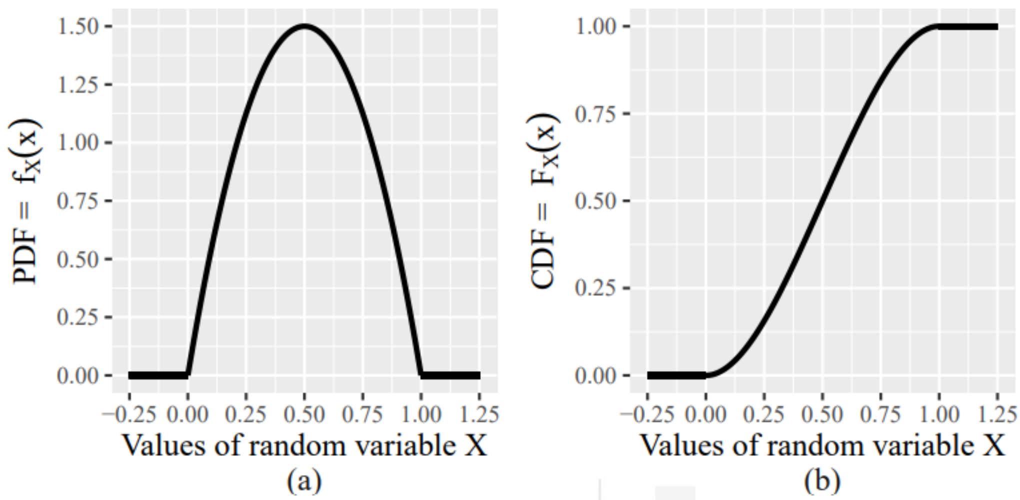

Fig. 5.4: Visualisation of (a) PDF (b) CDF of the random variable X defined in Example 2

(c) As mentioned in (5.17), we will use CDF to obtain required probabilities instead of PDF. So, using (5.22) required probabilities are given by

(i) $\mathcal{P}\left(\frac{1}{4}, \frac{1}{2}\right) = \int_{1/4}^{1/2} f_X(x)\,dx$

$$= F_X(1/2) - F_X(1/4) \quad [\text{Using } (5.22)]$$

$$= \left(3\left(\frac{1}{2}\right)^2 - 2\left(\frac{1}{2}\right)^3\right) - \left(3\left(\frac{1}{4}\right)^2 - 2\left(\frac{1}{4}\right)^3\right) = \frac{4}{8} - \frac{10}{64} = \frac{22}{64} = \frac{11}{32}$$

(ii) $\mathcal{P}\left(\frac{1}{3}, 1\right) = \int_{1/3}^{1} f_X(x)\,dx$

$$= F_X(1) - F_X(1/3) \quad [\text{Using } (5.22)]$$

$$= (1) - \left(3\left(\frac{1}{3}\right)^2 - 2\left(\frac{1}{3}\right)^3\right) \quad [\text{Using } (5.37)]$$

$$= 1 - \frac{7}{27} = \frac{20}{27}$$

(iii) $\mathcal{P}(-3, 2) = \int_{-3}^{2} f_X(x)\,dx$

$$= F_X(2) - F_X(-3) \quad [\text{Using } (5.22)]$$

$$= (1) - (0) = 1 \quad [\text{Using } (5.37)]$$

(iv) $\mathcal{P}(0.1, 1.1) = \int_{0.1}^{1.1} f_X(x)\,dx$

$$= F_X(1.1) - F_X(0.1) \quad [\text{Using } (5.22)]$$

$$= (1) - \left(3(0.1)^2 - 2(0.1)^3\right) \quad [\text{Using } (5.37)]$$

$$= 1 - (0.03 - 0.002) = 1 - 0.028 = 0.972$$

Example 3: Verify that CDF obtained in Example 2 satisfies all the four conditions of a CDF mentioned in (4.52) to (4.55) in Unit 4 of this course.

Solution: CDF obtained in Example 2 is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 3x^2 - 2x^3, & \text{if } 0 \le x < 1 \\ 1, & \text{if } x \ge 1 \end{cases} \quad \ldots (5.38)$$

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) $F_X(-\infty) = \lim\limits_{x \to -\infty} F_X(x) = \lim\limits_{x \to -\infty} (0)$ $\quad$ [Using (5.38)]

$= 0$

(b) $F_X(\infty) = \lim\limits_{x \to \infty} F_X(x) = \lim\limits_{x \to \infty} (1)$ $\quad$ [Using (5.38)]

$= 1$

(c) Let $x, y \in \mathbb{R}$ be such that $x \le y.$ From the graph of the function F_X shown in Fig. 5.4 (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 < 0 then using (5.38), we have $F_X(x) = 0, F_X(y) = 0,$ so $F_X(x) = F_X(y).$

Similarly, if both x and y are ≥ 1 then using (5.38), we have $F_X(x) = 1, F_X(y) = 1,$ so $F_X(x) = F_X(y).$

Finally, suppose $x, y \in [0, 1),$ i.e., $0 \le x, y < 1$, then

$$\begin{aligned} F_X(y) - F_X(x) &= \int_{-\infty}^{y} f_X(x)\,dx - \int_{-\infty}^{x} f_X(x)\,dx \\ &= \int_{-\infty}^{x} f_X(x)\,dx + \int_{x}^{y} f_X(x)\,dx - \int_{-\infty}^{x} f_X(x)\,dx \quad \text{[Using (5.62) of MST-011]} \\ &= \int_{x}^{y} f_X(x)\,dx \ge 0 \quad \left[\because y \ge x, \text{ so } \mathcal{P}\big(X \in (x, y)\big) \ge 0\right] \end{aligned}$$

$\Rightarrow F_X(y) - F_X(x) \ge 0 \Rightarrow F_X(y) \ge F_X(x) \Rightarrow F_X(x) \le F_X(y)$

Hence, $F_X(x) \le F_X(y)$ in all the possible cases. So, F_X is increasing or non-decreasing function.

(d) Here F_X is a piece wise function (having three pieces) and constant in two pieces and polynomial in one piece. We know that both a constant and polynomial functions are 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 = 0 and 1 only.

Right continuity at $x = 0$

$$\text{RHL}_{\text{at } x=0} = \lim_{x\to 0^+} F_X(x) = \lim_{x\to 0^+}(3x^2 - 2x^3) = 0, \text{ also } F_X(0) = 3(0)^2 - 2(0)^3 = 0.$$

Right continuity at $x = 1$

$$\text{RHL}_{\text{at } x=1} = \lim_{x\to 1^+} F_X(x) = \lim_{x\to 1^+}(1) = 1, \text{ also } F_X(1) = 3(1)^2 - 2(1)^3 = 3 - 2 = 1.$$

Since right hand limit (RHL) is equal to value of the function at that point. So, F_X is a right continuous function at x = 0 and 1.

Hence, F_X satisfies all the four conditions of CDF mentioned in (4.52) to (4.55) in Unit 4 of this course. This completes the solution of Example 2.

Recall first definition of continuous random variable given in (5.2) which says that CDF of a continuous random variable is differentiable. In Example 3, we have proved only right continuity of CDF. So, in the next example let us discuss continuity and differentiability of CDF of Example 3.

Example 4: Verify that CDF obtained in Example 2 is continuous and differentiable.

Solution: We know that if a function is differentiable at a point in its domain, then it is also continuous at that point. So, it is enough to verify differentiability of CDF obtained in Example 2 which is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 3x^2 - 2x^3, & \text{if } 0 \le x < 1 \\ 1, & \text{if } x \ge 1 \end{cases} \qquad \dots (5.39)$$

Here F_X is a piece wise function (having three pieces) and constant in two pieces and polynomial in one piece. We know that both a constant and polynomial functions are differentiable. So, to check that F_X is a differentiable function at each real number we have to check its differentiability only at the points which separate two pieces. Keeping this in view, we have to check differentiability of F_X at the points x = 0 and 1 only. … (5.40)

Differentiability at $x = 0$

$$\text{RHD}_{\text{at } x=0} = \lim_{h\to 0^+} \frac{F_X(0+h) - F_X(0)}{h} = \lim_{h\to 0^+} \frac{F_X(h) - 0}{h} = \lim_{h\to 0^+} \frac{3h^2 - 2h^3}{h}$$

$$= \lim_{h\to 0^+}(3h - 2h^2) = 0 \qquad \dots (5.41)$$

$$\text{LHD}_{\text{at } x=0} = \lim_{h\to 0^-} \frac{F_X(0+h) - F_X(0)}{h} = \lim_{h\to 0^-} \frac{F_X(h) - 0}{h} = \lim_{h\to 0^-} \frac{F_X(h)}{h}$$

$$= \lim_{h\to 0^-} \frac{0}{h} \qquad \left[\begin{array}{l} \because h \to 0^- \Rightarrow h \text{ is slightly less than 0 and} \\ \text{so by (5.39), we have } F_X(h) = 0 \end{array}\right]$$

$$= \lim_{h\to 0^-}(0) = 0 \qquad \dots (5.42)$$

From (5.41) and (5.42), we have $\text{LHD}_{\text{at } x=0} = \text{RHD}_{\text{at } x=0} = 0.$ Hence, CDF is differentiable at x = 0. … (5.43)

Differentiability at $x = 1$

$$\text{RHD}_{\text{at } x=1} = \lim_{h\to 0^+} \frac{F_X(1+h) - F_X(1)}{h} = \lim_{h\to 0^+} \frac{F_X(1+h) - 1}{h}$$

$$= \lim_{h\to 0^+} \frac{3(1+h)^2 - 2(1+h)^3 - 1}{h} \left[\begin{array}{l} \because h \to 0^+ \Rightarrow h \text{ is slightly greater than } 0 \\ \text{and so } 1+h \text{ is slightly greater than } 1 \\ \text{and using (5.39)} \end{array}\right]$$

$$= \lim_{h\to 0^+} \frac{-2h^3 - 3h^2}{h} = \lim_{h\to 0^+} \left(-2h^2 - 3h\right) = 0 \quad \text{... (5.44)}$$

$$\text{LHD}_{\text{at } x=1} = \lim_{h\to 0^-} \frac{F_X(1+h) - F_X(1)}{h} = \lim_{h\to 0^-} \frac{F_X(1+h) - 1}{h}$$

$$= \lim_{h\to 0^-} \frac{3(1+h)^2 - 2(1+h)^3 - 1}{h} \left[\text{Similar reason}\right]$$

$$= \lim_{h\to 0^-} \frac{-2h^3 - 3h^2}{h} = \lim_{h\to 0^-} \left(-2h^2 - 3h\right) = 0 \quad \text{... (5.45)}$$

From (5.44) and (5.45), we have $\text{LHD}_{\text{at } x=1} = \text{RHD}_{\text{at } x=1} = 0.$ Hence, CDF is differentiable at x = 1. … (5.46

On combining (5.40), (5.43) and (5.46), we can say that CDF obtain in Example 2 is a differentiable function. Thus, CDF is differentiable function and its derivative is given by

$$F_X'(x) = \begin{cases} 0, & \text{if } x < 0 \\ 6x - 6x^2, & \text{if } 0 \le x < 1 \\ 0, & \text{if } x \ge 1 \end{cases}$$

or $F_X'(x) = \begin{cases} 6x(1-x), & \text{if } 0 \le x < 1 \\ 0, & \text{otherwise} \end{cases}$

Or we can write it as follows.

$$\left.\begin{array}{l} F_X'(x) = f_X(x) \\ \text{i.e., Derivative of CDF} = \text{PDF} \\ \text{or antiderivative of PDF} = \text{CDF} \end{array}\right\} \quad \text{… (5.47)}$$

Remark 3: It did not happen by chance. This relationship holds in general between CDF and PDF of every continuous random variable. This can be easily obtained by combining (5.80) of MST-011 and (5.15) of this unit.

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

SAQ 2

For given CDF obtain PDF and probabilities: CDF of a random variable X is given by

$$F_X(x) = \begin{cases} 1 - e^{-2x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

(a) Obtain PDF of X and plot its graph.

(b) Plot the graph of CDF of X.

(c) Find probability that random variable X satisfies (i) X < 0.5, and (ii) X > 1.

5.4 SUMMARY

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

- **Continuous Variable**: If it is not possible to write all possible values of a variable in a sequence then such a variable falls in the category of continuous variable. In other words, we can say that list of possible values of a continuous variable is **uncountable** in nature.
- **First Definition of Continuous Random Variable**: A random variable X falls in the category of continuous random variable if its CDF is differentiable.
- **Second Definition of Continuous Random Variable**: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. A random variable X is said to be a continuous random variable if

$\mathcal{P}_\lambda(B) = 0 \ \forall \ B \in \mathcal{B}(\mathbb{R})$ whenever $\lambda(B) = 0$, where λ is a Lebesgue measure.

- **Cumulative Distribution Function of a Continuous Random Variable**: Let X be a continuous random variable then CDF of X is denoted by F_X and defined by $F_X(x) = \mathcal{P}(X \le x)$. That is like the discrete case CDF gives accumulated probability up to and including x.
- **Probability Density Function of a Continuous Random Variable**: Let X be a continuous random variable. The random variable X will have a probability density function (PDF) f_X if there exists a Borel measurable function $f_X : \mathbb{R} \to [0, \infty]$ such that

  $\mathcal{P}_X(B) = \mathcal{P}(X \in B) = \int_B f_X(x)\,dx, \ B \in \mathcal{B}(\mathbb{R}).$

- **Valid PDF**: The function f_X is called a valid PDF if it satisfies the following two conditions.
- $f_X(x) \ge 0 \ \forall \ x \in \mathbb{R}$ $[\because$ Probabilities are always non-negative$]$
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ $\left[\begin{array}{l}\because \text{Sum of all probabilities should be 1 and in} \\ \text{continuous world sum is given by integration}\end{array}\right]$
- **Interpretation of PDF**: Probability density function at a point X = x can be interpretated as **probability per unit length around the point X = x**.

5.5 TERMINAL QUESTIONS

1. **For given PDF obtain CDF and probabilities**: Revisit Example 2: Let X be the random variable having PDF as defined in Example 2. Obtain probability that X satisfies $|X| \le 1/2$.

2. **Calculation of Median and Percentile**: Revisit SAQ 2: Let X be the random variable having CDF as defined in SAQ 2.

   (a) Find x such that $\mathcal{P}(X \le a) = 0.2$

   (b) median of X, and (c) 90^{th} percentile of X.

3. **For given CDF obtain PDF and probabilities**: CDF of a random variable X is given by.

$$F_X(x) = \begin{cases} 1 - \dfrac{25}{x^2}, & x > 5 \\ 0, & \text{otherwise} \end{cases}$$

(a) Obtain PDF of X. Find probability that X satisfies (b) $X \le 6$ (c) X >7.

(d) Plot the graph of PDF and CDF of X.

5.6 SOLUTIONS/ANSWERS

Self-Assessment Questions (SAQs)

1. To understand the problem suppose weights vary from 495 ml to 505 ml. We are assuming that minimum weight is 495 ml exactly. Now, think can you write what is the next possible value of weight after 495 ml, if you write 495.1 ml, no between 495 ml and 495.1 ml we have possibility of 495.01 ml. If you say then after 495 ml next possible value is 495.01 ml. Again, the same issue between 495 ml and 495.01 ml, we have possibility of 495.001 ml. If you say then after 495 ml the next possible value is 495.001 ml. Again, the same issue between 495 ml and 495.001 ml we have possibility of 495.0001 ml. Continue like this you cannot write the next possible value after 495 ml. Hence, by nature this weight variable falls in the category of the continuous variables.

2. Given CDF of the random variable X is

$$F_X(x) = \begin{cases} 1 - e^{-2x}, & x > 0 \\ 0, & \text{otherwise} \end{cases} \qquad \dots (5.48)$$

(a) We can find PDF of X using (5.47) as follows.

$$\text{PDF} = f_X(x) = F_X'(x) = \begin{cases} 2e^{-2x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Graph of PDF is shown in Fig. 5.5 (a).

(b) Graph of CDF of X is shown in Fig. 5.5 (b).

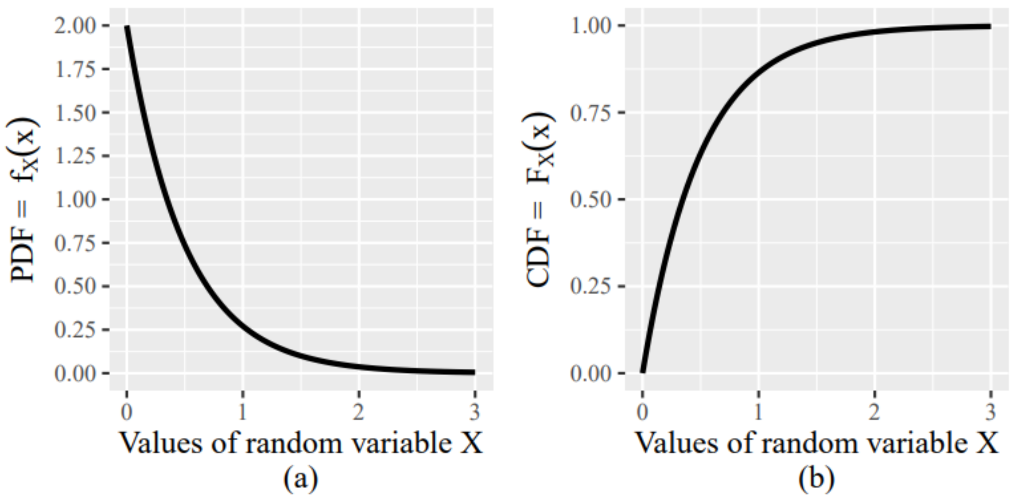

Fig. 5.5: Visualisation of (a) PDF (b) CDF of the random variable X defined in SAQ 2

(c) As mentioned in (5.17), we will use CDF to obtain required probabilities instead of PDF. So, using (5.22) required probabilities are given by

(i) $\mathcal{P}(X < 1/5) = \mathcal{P}(X \le 1/5) - \mathcal{P}(X = 1/5) = \mathcal{P}(X \le 1/5) - 0$ [Using (5.6)]

$= F_X(1/5)$ [Using (5.22)]

$= \left(1 - e^{-2(1/5)}\right) = 0.32968$

(ii) $\mathcal{P}(X > 1) = 1 - \mathcal{P}(X \le 1) = 1 - F_X(1)$ [Using (5.6)]

$= F_X(1)$ [Using (5.15)]

$= \left(1 - e^{-2(1)}\right) = 0.8646647$

Terminal Questions

1. In Example 2, we have obtained CDF of X which is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 3x^2 - 2x^3, & \text{if } 0 \le x < 1 \\ 1, & \text{if } x \ge 1 \end{cases} \quad \ldots (5.49)$$

Now, required probability is given by

$\mathcal{P}(|X| \le 1/3) = \mathcal{P}(-1/3 \le X \le 1/3)$ [Using (6.93) of MST-011]

$= F_X(1/3) - F_X(-1/3)$ [Using (5.22) of this unit]

$= \left(3(1/3)^2 - 2(1/3)^3\right) - (0)$ [Using (5.49) of this unit]

$= \frac{3}{9} - \frac{2}{27} = \frac{7}{27}$

2. In SAQ 2, given CDF of X is

$$F_X(x) = \begin{cases} 1 - e^{-2x}, & x > 0 \\ 0, & \text{otherwise} \end{cases} \quad \ldots (5.50)$$

(a) We have to find value of 'a' such that

$\mathcal{P}(X \le a) = 0.2$

$\Rightarrow F_X(a) = 0.2$ [Using (5.8)]

$\Rightarrow \left(1 - e^{-2a}\right) = 0.2$ [Using (5.49)]

$\Rightarrow e^{-2a} = 1 - 0.2 \Rightarrow e^{-2a} = 0.8 \Rightarrow -2a = \log 0.8 \Rightarrow a = -\frac{1}{2}(-0.2231436)$

$\Rightarrow a = 0.1115718$

(b) Let M be the value of the median of the random variable X, so, we must have

$\mathcal{P}(X \le M) = 0.5$

$\Rightarrow F_X(M) = 0.5$ [Using (5.8)]

$\Rightarrow \left(1 - e^{-2M}\right) = 0.5$ [Using (5.49)]

$\Rightarrow e^{-2M} = 1 - 0.5 \Rightarrow e^{-2M} = 0.5 \Rightarrow -2M = \log 0.5 \Rightarrow M = -\frac{1}{2}(-0.6931472)$

$\Rightarrow M = 0.3465736.$ So, median of X is 0.3465736.

(c) Let P be the value of the 90^{th} percentile of the random variable X, so, we must have

$$\mathcal{P}(X \le P) = 0.9$$
$$\Rightarrow F_X(P) = 0.9 \qquad [\text{Using } (5.8)]$$
$$\Rightarrow \left(1 - e^{-2P}\right) = 0.9 \qquad [\text{Using } (5.49)]$$
$$\Rightarrow e^{-2P} = 1 - 0.9 \Rightarrow e^{-2P} = 0.1 \Rightarrow -2P = \log 0.1 \Rightarrow P = -\frac{1}{2}(-2.302585)$$
$$\Rightarrow P = 1.1512925. \text{ So, } 90^{\text{th}} \text{ percentile of } X \text{ is } 1.1512925.$$

3. Given CDF of the random variable X is

$$F_X(x) = \begin{cases} 1 - \dfrac{25}{x^2}, & x > 5 \\ 0, & \text{otherwise} \end{cases} \qquad \ldots (5.51)$$

(a) Using (5.47) PDF of X is given by

$$\text{PDF} = f_X(x) = F_X'(x) = \begin{cases} \dfrac{50}{x^3}, & x > 5 \\ 0, & \text{otherwise} \end{cases}$$

(b) Now, required probability is given by

$$\mathcal{P}(X \le 6) = F_X(6) \qquad [\text{Using } (5.8)]$$
$$= 1 - \frac{25}{36} \qquad [\text{Using } (5.51) \text{ of this unit}]$$
$$= \frac{11}{36}$$

(c) Required probability is given by

$$\mathcal{P}(X > 7) = 1 - \mathcal{P}(X \le 7) = 1 - F_X(7) \qquad [\text{Using } (5.8)]$$
$$= 1 - \frac{25}{49} \qquad [\text{Using } (5.51)]$$
$$= \frac{24}{49}$$

(d) Graph of PDF and CDF are shown in Figs. 5.6 (a) and CDF of X is shown in Fig. 5.6 (b).

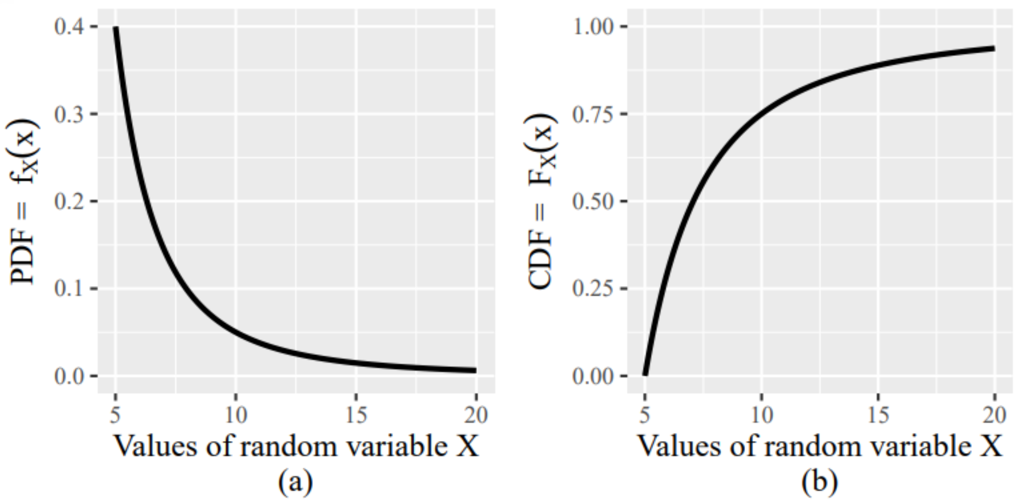

Fig. 5.6: Visualisation of (a) PDF (b) CDF of the random variable X defined in TQ 3