Chapter 2

Random Variables and Probability Distributions

Prof. N. G. J. Dias

Senior Professor and Head

Department of Computer Systems Engineering

Faculty of Computing and Technology

University of Kelaniya

2.1 Random Variables

2.1.1 Introduction to random Variables

- Let S be a sample space of an experiment.

  As noted previously, the outcome of the experiment, or the points in S, need not be numbers.

  For example, in tossing a coin the outcomes are H (heads) or T (tails), and in tossing a pair of dice the outcomes are pairs of integers.

- However, we frequently wish to assign a specific number to each outcome of the experiment.

  For example, in coin tossing, it may be convenient to assign 1 to H and 0 to T; or, in the tossing of a pair of dice, we may want to assign the sum of the two integers to the outcome.

- Such an assignment of numerical values is called a *random variable*.

  More generally, we have the following definition.

2.1.2 Definition of a Random Variable

- Definition 2.1

  A *random variable* X is a rule (function) that assigns (associates) a numerical value (a real number) to each outcome in a sample space S of the experiment.

  We shall let R_X denote the set of numbers assigned by a random variable X, and we shall refer to R_X as the *range space*.

  Clearly, a random variable is a function from the sample space S of the experiment to the real numbers $\mathbb{R}$ and R_X is the range of X.

- We shall use a capital letter, say X, to denote a random variable and its corresponding small letter, x in this case, for one of its values.

Example 2.1

Consider the experiment of tossing a pair of fair dice.

The sample space S consists of the 36 ordered pairs (a, b), where a and b can be any of the integers from 1 to 6: that is,

$$S = \{(1,1), \cdots, (1,6), (2,1), \cdots, (2,6), (3,1), \cdots, (3,6), (4,1), \cdots, (4,6),$$
$$(5,1), \cdots, (5,6), (6,1), \cdots, (6,6)\}.$$

Let X assign to each point in S the sum of the numbers; then X is a random variable with range space $R_X = \{2,3,4,5,6,7,8,9,10,11,12\}$.

Let Y assign to each point the maximum of the two numbers; then Y is a random variable with range space $R_Y = \{1,2,3,4,5,6\}$.

Note that the random variable X assumes the value 8 for all elements in the subset $E = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$ of the sample space S.

That is, each possible value of X represents an event that is a subset of the sample space for the given experiment.

Example 2.2

The sample space of each possible outcome when three electronic components are tested may be written

$$S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\},$$

where N denotes nondefective and D denotes defective.

Let X be the the number of defectives that occur.

Thus, each point in the sample space will be assigned a numerical value of 0, 1, 2, or 3.

Then X is a random variable with range space $R_X = \{0,1,2,3\}$.

Note that the random variable X assumes the value 2 for all elements in the subset $E = \{DDN, DND, NDD\}$ of the sample space S.

That is, each possible value of X represents an event that is a subset of the sample space for the given experiment.

Example 2.3

A box contains 12 items of which 3 are defective. A sample of three items is selected from the box.

The sample space S consists of $\binom{12}{3} = 220$ different samples of size 3.

Let X denote the number of defective items in the sample; then X is a random variable with range space $R_X = \{0,1,2,3\}$.

Example 2.4

Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls.

For this experiment sample space $S = \{RR, RB, BR, BB\}$.

Let Y be the number of red balls drawn.

The possible outcomes and the values y of the random variable Y, are given in the following table:

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Example 2.5

Statisticians use sampling plans to either accept or reject batches or lots of material.

Suppose one of these sampling plans involves sampling independently 10 items from a lot of 100 items in which 12 are defective.

In this experiment the sample space consists of $\binom{100}{10}$ different samples of size 10.

Let X be the random variable defined as the number of items found defective in the sample of 10.

In this case, the random variable X takes on the values 0, 1, 2, . . . , 9, 10.

Example 2.6

Let X be the random variable defined by the waiting time, in hours, between successive speeders spotted by a radar unit.

The random variable X takes on all values x for which $x \geq 0$.

Example 2.7

Consider the experiment of tossing a coin repeatedly until a head appears.

Let X be the random variable that the number of tosses required to get the first head, and

Y be the random variable that the number of tosses required before getting the first head.

Then the sample space for the experiment $S = \{\mathrm{H}, \mathrm{TH}, \mathrm{TTH}, \mathrm{TTTH}, \cdots\}$ and has infinite number of elements.

Thus, $R_X = \{1,2,3,\cdots\}$ and $R_Y = \{0,1,2,3,\cdots\}$.

2.1.3 Discrete and Continuous Random Variables

- A random variable is called a *discrete random variable* if its set of possible outcomes is countable.
- In other words, random variables whose set of possible values can be written either as a finite sequence $x_1, x_2, \cdots, x_n$, or as an infinite sequence $x_1, x_2, \cdots$ are said to be *discrete*.
- The random variables in Examples 2.1 to 2.4 are discrete random variables.
- There are random variables whose set of possible values is an entire interval of numbers on the real line, and hence are not discrete.
- When a random variable can take on values on a continuous scale (uncountable), it is called a *continuous random variable*.
- Often the possible values of a continuous random variable are precisely the same values that are contained in the continuous sample space.

- Obviously, the random variable described in Example 2.5 is a continuous random variable.
- Another example for a continuous random variable is the random variable denoting the lifetime of a car, when the car's lifetime is assumed to take on any value in some interval (a, b).
- In most practical problems, continuous random variables represent *measured* data, such as all possible heights, weights, temperatures, distance, or life periods, whereas discrete random variables represent *count* data, such as the number of defectives in a sample of k items or the number of highway fatalities per year in a given city.

Exercises

1. Classify the following random variables as discrete or continuous:

    X: the number of automobile accidents per year in Virginia.

    Y: the length of time to play 18 holes of golf.

    M: the amount of milk produced yearly by a particular cow.

    N: the number of eggs laid each month by a hen.

    P: the number of building permits issued each month in a certain city.

    Q: the weight of grain produced per acre.

2. An overseas shipment of 5 foreign automobiles contains 2 that have slight paint blemishes. If an agency receives 3 of these automobiles at random, list the elements of the sample space S, using the letters B and N for blemished and nonblemished, respectively; then to each sample point assign a value x of the random variable X representing the number of automobiles with paint blemishes purchased by the agency.

2.2 Discrete Probability Distributions

2.2.1 Probability Mass Function (pmf)

- A discrete random variable assumes each of its values with a certain probability.
- Let us consider the case of tossing a coin three times and let X be the random variable representing the number of heads.

  For this experiment the sample space S consists of 8 elements and is given by

  $$S = \{TTT, HTT, THT, TTH, HHT, HTH, THH, HHH\}.$$

  The range space of the random variable X is $R_X = \{0,1,2,3\}$.

  X assumes the value 0 with probability 1/8, since only 1 of the 8 equally likely sample points result in zero heads and three tails.

  X assumes the value 1 with probability 3/8, since 3 of the 8 equally likely sample points result in one head and two tails.

  X assumes the value 2 with probability 3/8, since 3 of the 8 equally likely sample points result in two heads and one tail.

X assumes the value 3 with probability 1/8, since only 1 of the 8 equally likely sample points result in 3 heads and zero tails.

Now we can summarize the values x of the random variable X together with the corresponding probabilities as in in the following table:

x	0	1	2	3
$P(X = x)$	$^1/_8$	$^3/_8$	$^3/_8$	$^1/_8$

Notice that, $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1$ and

$P(X = x) > 0$ for all x in R_X.

- Frequently, it is convenient to represent all the probabilities of a random variable X by a formula.
- Such a formula would necessarily be a function of the numerical values x that we shall denote by $f(x)$ (or $f_X(x)$).

- Definition 5 - Probability Mass Function of a Discrete Random Variable

  Let X be a discrete random variable with distinct values $x_1, x_2, \cdots, x_n, \cdots$.

  Then the function, denoted by $f(x)$ (or $f_X(x)$) and defined by

$$f(x) = \begin{cases} P(X = x_i), & \text{if } x = x_i, i = 1,2,\cdots,n,\cdots \\ 0, & \text{if } x \neq x_i, \end{cases}$$

  is called the *probability function* or *probability mass function* of the discrete random variable X.

  The set of ordered pairs $(x_1, f(x_1)), (x_2, f(x_2)), \cdots, (x_n, f(x_n)), \cdots$, is called the *probability distribution* of the discrete random variable X, and it is usually given as a formula or as a table.

- Note that, $f(x)$ (or $f_X(x)$) is a function with domain the real line and codomain the interval $[0,1]$.
- A graph of $f(x)$ (or $f_X(x)$) is called a *probability graph*.

- In general, a *probability function* or *probability mass function* of a discrete random variable X satisfy the following two properties:

  1. $f(x) \geq 0,$
  2. $\sum_x f(x) = 1,$

  for each possible outcome x of the discrete random variable X.

- In the case that, the sample space S is an equiprobable space, we can easily obtain the distribution of a random variable from the following result.

- Theorem 2.1

  Let S be an equiprobable space, and let X be a random variable on S with range space $R_X = \{x_1, x_2, \cdots, x_n\}$.

  Then $f(x_i) = P(X = x_i) = \dfrac{\text{number of points in } S \text{ whose image is } x_i}{\text{number of points in } S}$

Example 2.8

Again consider the experiment of tossing of a pair of fair dice in Example 2.1.

Let X be the random variable which gives the sum of the numbers appear on the two dice.

There are 36 outcomes in the sample space and range space $R_X = \{2,3,\cdots,12\}$.

There is only one outcome $(1,1)$ whose sum is 2; hence $P(X = 2) = \frac{1}{36}$.

There are two outcomes, $(1,2)$ and $(2,1)$, whose sum is 3; hence $P(X = 3) = \frac{2}{36}$.

There are three outcomes, $(1,3), (2,2)$ and $(3,1)$, whose sum is 4; hence $P(X = 4) = \frac{3}{36}$.

Similarly, $P(X = 5) = \frac{4}{36}, P(X = 6) = \frac{5}{36}, \cdots, P(X = 12) = \frac{1}{36}$.

The probability distribution of X consists of the points in R_X along with their respective probabilities given by the following table:

x_i	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_i\}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Example 2.9

Again consider the experiment of selecting a sample of 3 items from a box containing 12 items of which 3 are defective in Example 2.3.

The sample space S consists of $\binom{12}{3} = 220$ different samples of size 3.

Let X denote the number of defective items in the sample; then X is a random variable with range space $R_X = \{0,1,2,3\}$.

There are $\binom{9}{3} = 84$ samples of size 3 with no defective items; hence $P(x =$

There is only one sample of size 3 containing the three defective items; hence $P(x = 3) = \frac{1}{220}$.

Then the probability distribution of X is as follows:

x_i	0	1	2	3
$P(X = x_i)$	$\frac{84}{220}$	$\frac{108}{220}$	$\frac{27}{220}$	$\frac{1}{220}$

Example 2.10

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school.

Then x can only take the numbers 0, 1, and 2.

Now $f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}$, $f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$,

and $f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$.

Thus the probability distribution of X is

x_i	0	1	2
$P(X = x_i)$	$68/95$	$51/190$	$3/190$

2.2.2 Cumulative Distribution Function (cdf)

- There are many problems where we may wish to compute the probability that the observed value of a random variable X will be less than or equal to some real number x.
- By writing $F(x) = P(X \leq x)$ for every real number x, we define $F(x)$ (or $F_X(x)$) to be the *cumulative distribution function* of the random variable X.

- Definition 6 - Cumulative Distribution Function of a Discrete Random Variable

The *cumulative distribution function* $F(x)$ (or $F_X(x)$) of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t),$$

where x is any real number, i.e., $-\infty < x < \infty$.

- The cumulative distribution function $F(x)$ has the following properties:
  1. $F(x)$ is nondecreasing [i.e., $F(x) \leq F(y)$ if $x \leq y$].
  2. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.
  3. $F(x)$ is continuous from the right [i.e., $\lim_{h \to 0^+} F(x+h) = F(x)$ for all x].
- Suppose that the discrete random variable X takes on only a finite number of values $x_1, x_2, \cdots, x_n$. Then the cumulative distribution function $F(x)$ is given by

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3. \\ \vdots & \vdots \\ f(x_1) + f(x_2) + \cdots + f(x_n) & x_n \leq x < \infty \end{cases}$$

Example 2.11

Suppose that a fair coin is tossed twice so that the sample space is $S = \{HH, HT, TH, TT\}$.

Let X represent the number of heads that can come up.

Thus, range space $R_X = \{0,1,2\}$.

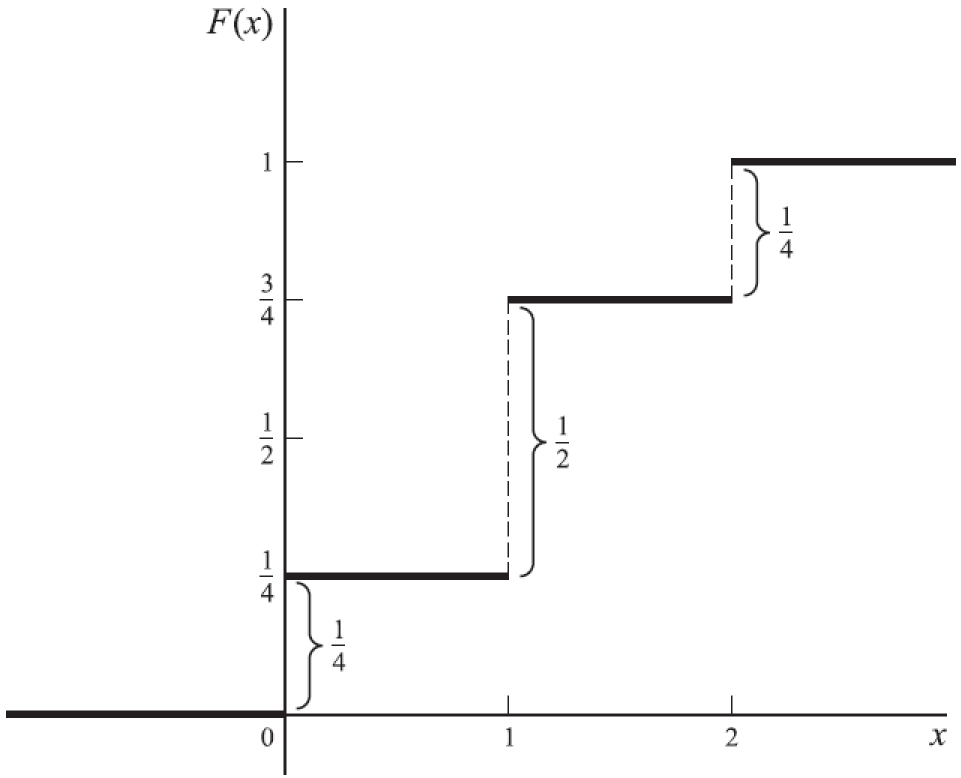
The probability density function corresponding to the random variable X is given by

x_i	0	1	2
$P(X = x_i)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

The cumulative distribution function is given by

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ 1/4 & 0 \le x < 1 \\ 3/4 & 1 \le x < 2 \\ 1 & 2 \le x < \infty \end{cases}.$$

The graph of $F(x)$ is given below:



- One should pay particular notice to the fact that the cumulative distribution function is a monotone non-decreasing function defined not only for the values assumed by the given random variable but for all real numbers.

Example 2.12

Again consider the experiment of tossing of a pair of fair dice discussed in Example 2.1 and Example 2.8.

Let X be the random variable which gives the sum of the numbers appear on the two dice.

As you know, there are 36 outcomes in the sample space and range space $R_X = \{2,3,\cdots,12\}$.

From Example 2.8, the probability distribution of the random variable X is given by the following table:

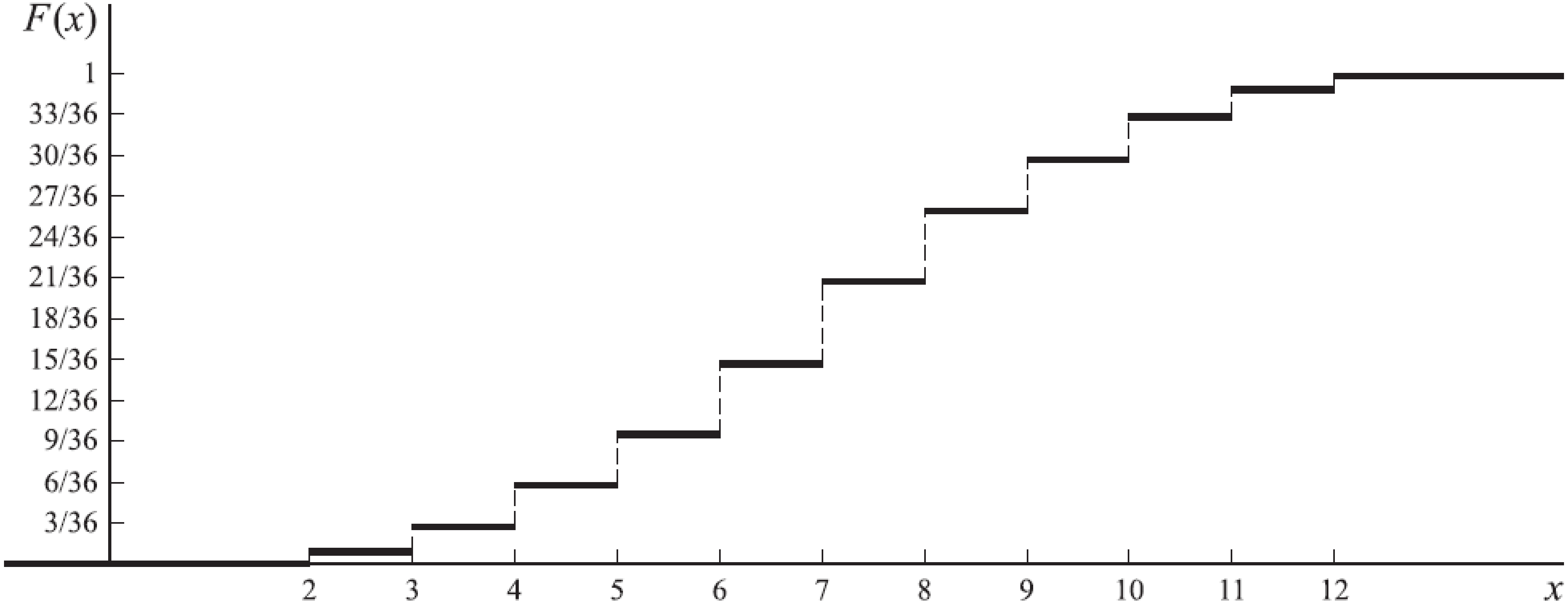
x_i	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_i\}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

By using $F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$ and the probability distribution table we can compute the cumulative distribution function of the random variable X.

Thus,

$$F(x) = \begin{cases} 0, & -\infty < x < 2 \\ 1/36, & 2 \leq x < 3 \\ 3/36, & 3 \leq x < 4 \\ 6/36, & 4 \leq x < 5 \\ \vdots & \vdots \\ 35/36, & 11 \leq x < 12 \\ 1, & 12 \leq x < \infty \end{cases} .$$

The graph of the cumulative distribution function $F(x)$ is given below.



F(x)
1
33/36
30/36
27/36
24/36
21/36
18/36
15/36
12/36
9/36
6/36
3/36
2
3
4
5
6
7
8
9
10
11
12
x

Example 2.13

If a car agency sells 50% of its inventory of a certain foreign car equipped with side airbags, find a formula for the probability distribution of the number of cars with side airbags among the next 4 cars sold by the agency.

Hence, obtain the cumulative distribution function for the above probability distribution.

Plot the probability distribution and the cumulative distribution function.

Solution:

The $2^4 = 16$ points in the sample space are equally likely to occur, since the probability of selling an automobile with side airbags is $\frac{50}{100} = \frac{1}{2}$.

The event of selling x models with side airbags and $4 - x$ models without side airbags can occur in $\binom{4}{x}$ ways, where x can be 0, 1,2, 3, or 4.

Thus, the probability of selling x models with side airbags is given by

$$P(X = x) = \binom{4}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{4-x} = \binom{4}{x}\left(\frac{1}{2}\right)^{4} = \frac{1}{16}\binom{4}{x},$$

where x can be $0,1,2,3$ or 4.

Let be X be the random variable of selling x models with side airbags.
Then the probability distribution of the random variable X is given by

$$f(x) = P(X = x) = \frac{1}{16}\binom{4}{x}, \text{ where } x \text{ can be } 0,1,2,3 \text{ or } 4.$$

From $f(x) = \frac{1}{16}\binom{4}{x}$, we get $f(0) = \frac{1}{16}, f(1) = \frac{1}{4}, f(2) = \frac{3}{8}, f(3) = \frac{1}{4}$, and $f(4) = \frac{1}{16}$.

Therefore, $F(0) = f(0) = \frac{1}{16}$,

$$F(1) = f(0) + f(1) = \frac{1}{16} + \frac{1}{4} = \frac{5}{16},$$

$$F(2) = f(0) + f(1) + f(2) = \frac{5}{16} + \frac{3}{8} = \frac{11}{16},$$

$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{11}{16} + \frac{1}{4} = \frac{15}{16},$$

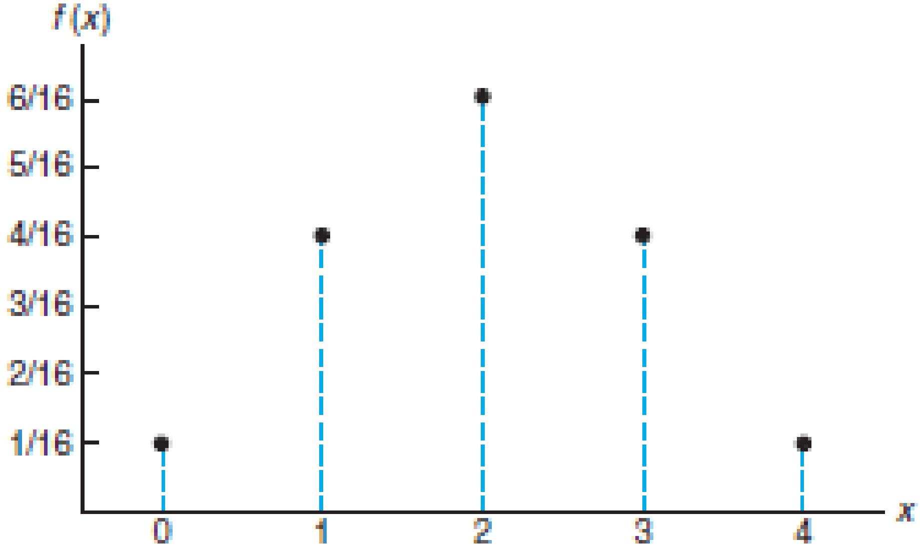
$$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = \frac{15}{16} + \frac{1}{16} = 1.$$

Hence,

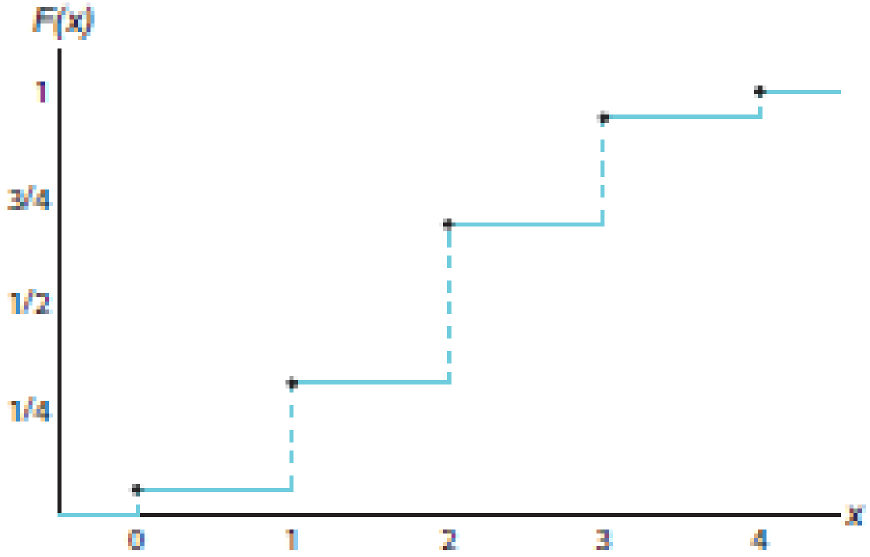
$$F(x) = \begin{cases} 0, & -\infty < x < 0, \\ \frac{1}{16}, & 0 \leq x < 1, \\ \frac{5}{16}, & 1 \leq x < 2, \\ \frac{11}{16}, & 2 \leq x < 3, \\ \frac{15}{16}, & 3 \leq x < 4, \\ 1, & 4 \leq x < \infty. \end{cases}.$$

Further we can observe that, $f(2) = F(2) - F(1) = \frac{11}{16} - \frac{5}{16} = \frac{6}{16} = \frac{3}{8}$.

Plots of the probability distribution and the cumulative distribution function are given below:



Plot of probability distribution



Plot of cumulative distribution function

Exercises

1. Determine the value c so that each of the following functions can serve as a probability distribution of the discrete random variable X:

   (a) $f(x) = c(x^2 + 4)$, for $x = 0,1,2,3$;

   (b) $f(x) = c\binom{2}{x}\binom{3}{3-x}$, for $x = 0,1,2.$

2. Find the probability distribution of boys and girls in families with 3 children, assuming equal probabilities for boys and girls.

   If X represents the number of boys in the family, find the sample space S, range space R_X and the probability distribution of X.

   Find the cumulative distribution function $F(x)$ for the random variable X and plot this cumulative distribution function $F(x)$.

3. A shipment of 7 television sets contains 2 defective sets. A hotel makes a random purchase of 3 of the sets.

   If x is the number of defective sets purchased by the hotel, find the probability distribution of X.

   Find the cumulative distribution function of the random variable X representing the number of defectives

   Then using the cumulative distribution function $F(X)$, find

   (a) $P(X = 1)$;

   (b) $P(0 < X \leq 2)$.

2.3 Continuous Probability Distributions

2.3.1 Probability Density Function (pdf)

- A continuous random variable has a probability of 0 of assuming *exactly* any of its values.
- Consequently, its probability distribution cannot be given in tabular form.
- In this case we are dealing with an interval rather than a point value of our random variable.
- In order to arrive at a probability distribution for a continuous random variable X, we note that the probability that X lies between two different values is meaningful.

- Definition 7 - Probability Density Function of a Continuous Random Variable

  The function $f(x)$ (or $f_X(x)$) is a *probability density function* (pdf) of a continuous random variable X, defined over the set of real numbers $\mathbb{R}$, if:

  1. $f(x) \geq 0$ for all $x \in \mathbb{R}$,
  2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$,
  3. $P(a < X < b) = \int_a^b f(x)\, dx$.

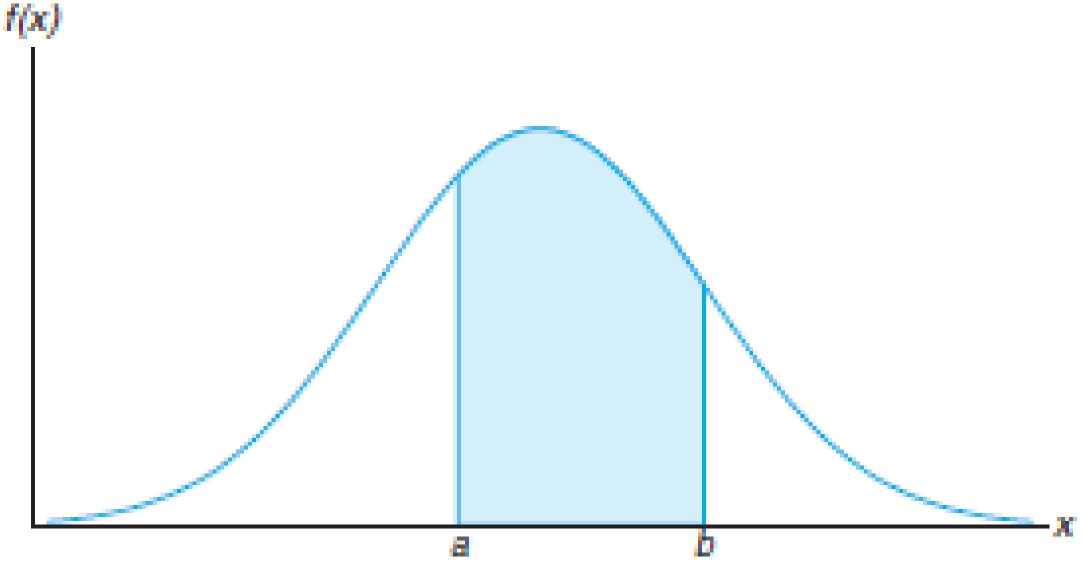
- Also note that, for any continuous random variable X,

$$P(a < X \leq b) = P(a < X < b) + P(X = b) = P(a < X < b).$$

  That is, it does not matter whether we include an endpoint of the interval or not.

- Note that the probability that X assumes a value between a and b is equal to the shaded area under the probability density function between the ordinates at $x = a$ and $x = b$, and from integral calculus this area is given by

$$P(a < X < b) = \int_a^b f(x)\,dx.$$



Example 2.14

- Suppose that the error in the reaction temperature, in $^{\circ}C$, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & otherwise. \end{cases}$$

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

Solution: We use Definition 7.

(a) Obviously, $f(x) \geq 0$. To verify condition 2 in Definition 7, we have

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-1}^{2} \frac{x^2}{3} dx = \left[\frac{x^3}{9}\right]_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1.$$

(b) Using formula 3 in Definition 7, we obtain

$$P(0 < X < 1) = \int_0^1 \frac{x^2}{3}\,dx = \left[\frac{x^3}{9}\right]_0^1 = \frac{1}{9}.$$

Example 2.15

Find the constant c such that the function

$$f(x) = \begin{cases} cx^2, & 0 < x < 3 \\ 0, & otherwise \end{cases}$$

is a probability density function, and compute $P(1 < X < 2)$.

Solution: Since $f(x)$ satisfies Property 1 if $c \geq 0$, it must satisfy Property 2 in order to be a probability density function.

Now, $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^3 cx^2\,dx = \left[\frac{cx^3}{3}\right]_0^3 = 9c$,

and since this must be equal to 1 we have $c = \frac{1}{9}$.

Using formula 3 in Definition 7, we obtain

$$P(1 < X < 2) = \int_1^2 \frac{1}{9} x^2 \, dx = \left[\frac{x^3}{27}\right]_1^2 = \frac{8}{27} - \frac{1}{27} = \frac{7}{27}.$$

Since X is a continuous random variable, also note that

$$P(1 \le X \le 2) = P(1 \le X < 2) = P(1 < X \le 2) = P(1 < X < 2) = \frac{7}{27}.$$

2.3.2 Cumulative Distribution Function (cdf)

- **Definition 8 - Cumulative Distribution Function of a Continuous Random Variable**

  The *cumulative distribution function* $F(x)$ (or $F_X(x)$) of a continuous random variable X with probability density function $f(x)$ is

  $$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt,$$

  where x is any real number, i.e., $-\infty < x < \infty$.

- As an immediate consequence of Definition 8, one can write the two results

  $$P(a < X < b) = F(b) - F(a) \text{ and } f(x) = \frac{dF(x)}{dx},$$

  if the derivative exists.

  Further note that, $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$ and $0 \leq F(x) \leq 1$.

Example 2.16

Again consider the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & otherwise. \end{cases}$$

defined in Example 2.14.

Find the cumulative distribution function $F(x)$ and use it to evaluate $P(0 < X \leq 1)$.

Solution: We have $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$.

If $x < -1$, then $F(x) = 0$.

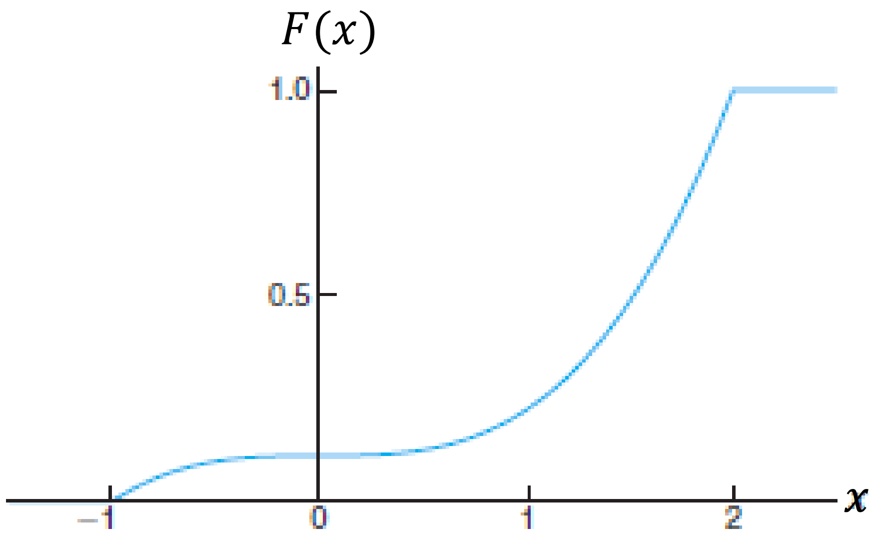
For $-1 < x < 2$, $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{x} \frac{t^2}{3}\,dt = \left[\frac{t^3}{9}\right]_{-1}^{x} = \frac{x^3+1}{9}$.

If $x \geq 2$, then $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{-1} f(t)\,dt + \int_{-1}^{2} f(t)\,dt + \int_{2}^{x} f(t)\,dt$

$$= \int_{-\infty}^{-1} 0\,dt + \int_{-1}^{2} \frac{t^2}{3}\,dt + \int_{2}^{x} 0\,dt$$

$$= \left[\frac{t^3}{9}\right]_{-1}^{2} = 1.$$

Therefore, $F(x) = \begin{cases} 0, & x < -1, \\ \frac{x^3+1}{9}, & -1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$



The cumulative distribution function $F(x)$ is given above.

Now, $P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$, which agrees with the result obtained by using the density function in Example 2.14.

Note that $F(x)$ increases monotonically from 0 to 1 as is required for a cumulative distribution function.

It should also be noted that $F(x)$ in this case is continuous.

Example 2.17

Again consider the probability density function

$$f(x) = \begin{cases} \frac{x^2}{9}, & 0 < x < 3, \\ 0, & otherwise. \end{cases}$$

defined in Example 2.15.

Find the cumulative distribution function $F(x)$ for the continuous random variable X and use it to evaluate $P(1 < X \leq 2)$.

Solution: We have $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$

If $x < 0$ then $F(x) = 0$.

For $-1 < x < 3$, $F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-\infty}^{x} \frac{t^2}{9} dt = \left[\frac{t^3}{27}\right]_{-1}^{x} = \frac{x^3}{27}$.

If $x \geq 3$ then $F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-\infty}^{0} f(t)\, dt + \int_{0}^{3} f(t)\, dt + \int_{3}^{x} f(t)\, dt$

$$= \int_{-\infty}^{0} 0\, dt + \int_{0}^{3} \frac{t^2}{9} dt + \int_{3}^{x} 0\, dt = \left[\frac{t^3}{27}\right]_{0}^{3} = 1.$$

Thus the required cumulative distribution function is

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{x^3}{27}, & 0 \le x < 3, \\ 1, & x \ge 3. \end{cases}$$

Note that $F(x)$ increases monotonically from 0 to 1 as is required for a cumulative distribution function.

It should also be noted that $F(x)$ in this case is continuous.

We have, $P(1 < X \le 2) = F(2) - F(1) = \frac{2^3}{27} - \frac{1^3}{27} = \frac{7}{27}$, which agrees with the result obtained by using the probability density function in Example 2.15.

Example 2.18 The Department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate b. The DOE has determined that the probability density function of the winning (low) bid is

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5b} \leq y \leq 2b, \\ 0, & elsewhere. \end{cases}$$

Find cumulative distribution function $F(y)$ and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate b.

Solution: For $\frac{2}{5b} \leq y \leq 2b$, $\quad F(y) = \int_{2/5b}^{y} \left(\frac{5}{8b}\right) dt = \left[\frac{5t}{8b}\right]_{2/5b}^{y} = \frac{5y}{8b} - \frac{1}{4}.$

$$\text{Thus, } F(y) = \begin{cases} 0, & y < \frac{2}{5b}, \\ \frac{5y}{8b} - \frac{1}{4}, & \frac{2}{5b} \leq y < 2b; \\ 1 & y \geq 2b \end{cases}$$

To determine the probability that the winning bid is less than the preliminary bid estimate b, we have

$$P(Y \leq b) = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}.$$

Exercises

1. The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the probability density function

$$f(x) = \begin{cases} x, & 0 < x < 1, \\ 2 - x, & 1 \leq x < 2, \\ 0, & otherwise. \end{cases}$$

   Find the probability that over a period of one year, a family runs their vacuum cleaner

   (a) less than 120 hours;

   (b) between 50 and 100 hours.

2. The proportion of people who respond to a certain mail-order solicitation is a continuous random variable X that has the probability density function

$$f(x) = \begin{cases} \frac{2(x+2)}{5}, & 0 < x < 1, \\ 0, & elsewhere. \end{cases}.$$

(a) Show that $P(0 < X < 1) = 1.$

(b) Find the probability that more than 1/4 but fewer than 1/2 of the people contacted will respond to this type of solicitation.

3. The waiting time, in hours, between successive speeders spotted by a radar unit is a continuous random variable with cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-8x}, & x \geq 0. \end{cases}.$$

Find the probability of waiting less than 12 minutes between successive speeders

(a) using the cumulative distribution function of X;

(b) using the probability density function of X.

4. A continuous random variable X that can assume values between $x = 1$ and $x = 3$ has a probability density function given by $f(x) = \frac{1}{2}$.

(a) Show that the area under the curve is equal to 1.

(b) Find $P(2 < X < 2.5)$.

(c) Find $P(X \leq 1.6)$.

Find the cumulative distribution function $F(x)$ of X.

5. A continuous random variable X that can assume values between $x = 2$ and $x = 5$ has a probability density function given by $f(x) = \frac{2(1+x)}{27}$.

   Find (a) $P(X < 4)$;

   (b) $P(3 \leq X < 4)$.

   Find the cumulative distribution function $F(x)$, and use it to evaluate $P(3 \leq X < 4)$.

6. Consider the probability density function $f(x) = \begin{cases} k\sqrt{x}, & 0 < x < 1, \\ 0, & otherwise. \end{cases}$.

   (a) Evaluate k.

   (b) Find the cumulative distribution function $F(x)$, and use it to evaluate $P(0.3 < X < 0.6)$.