Elements of

Probability

Contents

1 Introduction

Our daily experience in ordinary, in practical situations, and also in scientific investigations, constantly supply us with examples of the breakdown of the familiar regularities of strict determinism that we are accustomed to. For instance, suppose we wish to know how many telephone calls a first–aid station will receive within twenty four hours.

Long term observations indicate that there is no way of predicting the number of such calls. This number is subject to appreciable (and, what is more, random) fluctuations. But it not sufficient to merely indicate the fact of randomness in order to make use of a particular phenomena of nature or to control a technological process. We need to learn to evaluate random events numerically and predict the course they will take. Two divisions of mathematics are engaged in the solution of such problems and in constructing the requisite general mathematical theory: *theory of probability*, and *mathematical statistics*. In this handout, we deal with the basics of probability theory.

2 Preliminaries

2.1 Random Experiment. Sample Spaces

The basic idea behind the probability theory is that of *random experiment* or *random activity.*

▷ **Definition 1 (Random Experiment).** An activity is called a random experiment if all the possible *outcome* (or results) of the activity are known in advance, but none of which can be predicted with certainty.

***Example* 1.** The tossing of a (fair) coin is a random experiment because we know all the possible outcome: a head or a tail; but neither of these can be foretold.

▷ **Definition 2 (Sample Space).** The set of *all* possible outcomes of a random experiment is called the *sample space*, $\mathcal{S}$, associated with that (random) experiment and each individual outcome is called a *sample point.*

***Example* 2.** When we toss a fair coin, the sample space may be depicted as

$$\mathcal{S} = \{\text{Head},\ \text{Tail}\}.$$

***Example* 3.** A homogeneous cube with faces labeled with numbers from 1 to 6 (or marked with corresponding number of dots) is called a *die.* If we roll a die, then any one of the faces may show up. So the sample space in this case is the following set:

$$\mathcal{S} = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}.$$

Often, it is useful to depict the sample space in a diagram. To this end, we can assign numbers (whenever applicable) to each outcome and then depict them in a diagram.

***Example* 4.** Suppose two coins (one rupee coin and two rupee coin) are tossed once. Since we can easily distinguish the two coins, we can easily talk of coin number 1 ($C1$) and coin number 2 ($C2$). Also either coin can turn up with a head (H) or a tail (T). Thus the possible outcomes are:

(a) $(H,\ H) = HH =$ heads on both $C1$ and $C2$;

(b) $(H,\ T) = HT =$ head on $C1$, tail on $C2$;

(c) $(T,\ H) = TH =$ tail on $C1$, head on $C2$; and

(d) $(T,\ T) = TT =$ tails on both $C1$ and $C2$.

Hence the sample space $\mathcal{S} = \{HH,\ HT,\ TH,\ TT\}$. To depict the sample space in a diagram, let us assign the head a numerical value 0 and the tail 1. Then heads on both $C1$ and $C2$ can be assigned an ordered pair $(0,\ 0)$ with the understanding that the first of these numbers corresponds to $C1$ and the second to $C2$. Now we can depict the sample space of this experiment as shown in Fig. 1.

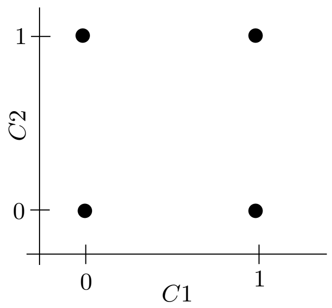

Figure 1: The sample space when two coins are tossed once.

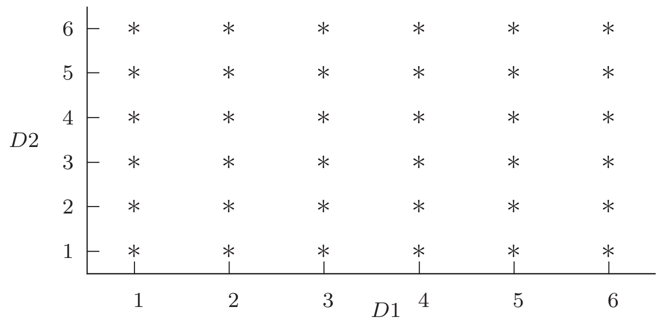

Figure 2: Sample space when a pair of dice is rolled once.

***Example* 5.** Let a pair of dice (plural of a die) be rolled once. Let us call them $D1$ and $D2$ respectively. In this case any number from 1 to 6 can be on $D1$ and similarly for $D2$. Let us associate an ordered pair $(i,\ j)$ to the occurrence of the number i on $D1$ and j on $D2$ where each of i and j can be any integer from 1 to 6. Then the sample space of this experiment becomes the set

$$\mathcal{S} = \{(i,\ j)|1 \leq i, j \leq 6;\ i, j \in \mathbf{I}\}.$$

This sample space has been depicted in Fig. 2.

***Example* 6.** From a group of 3 boys and 2 girls, we select two persons. What is the sample space?

There are only three possible cases: (i) both are boys; (ii) one is a boy and one is a girl; and (iii) both are girls. We can depict the situation (and the choices) as shown in Fig. 3. The diagram is called a *tree diagram* for obvious reasons. Thus the sample space:

$$\mathcal{S} = \{B_1B_2,\ B_1B_3,\ B_1G_1,\ B_1G_2,\ B_2B_3,\ B_2G_1,\ B_2G_2,\ B_3G_1,\ B_3G_2,\ G_1G_2\}.$$

The same can be also depicted as in Fig. 4. Alternatively, we can see that the number of ways in which we can select two persons out of five is given by ${}^5C_2 = 10$.

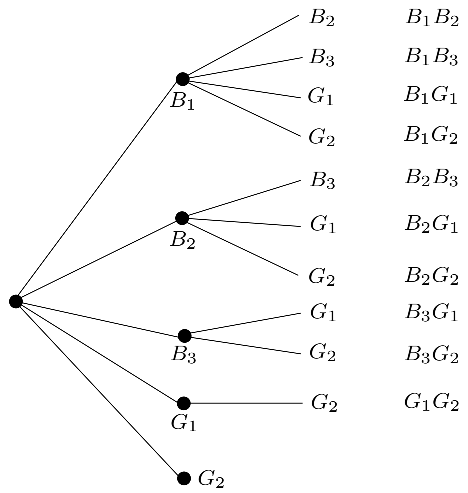

Figure 3: The tree diagram for determining the sample space in Example 6

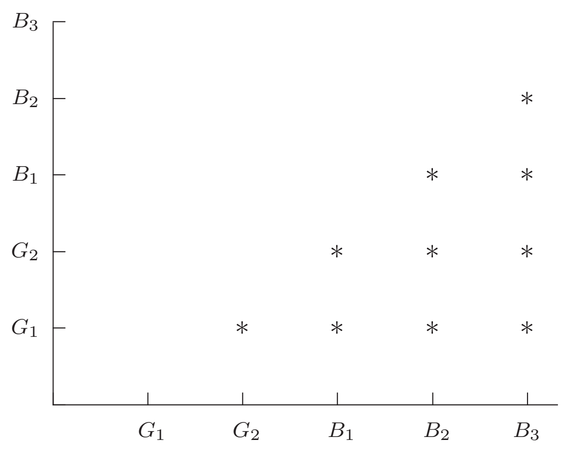

Figure 4: Sample space for Example 6.

♣ *Remark* 1. If a sample space consists of a finite number of sample points, it is called *finite sample space.* If it consists of as many sample points as there are natural numbers 1, 2, 3, ..., it is called *countably infinite sample space.* If it consists of as many sample points as there are in an interval, e.g. $0 \le x \le 1$, it is called *non–countably infinite sample space.* Finite and countably infinite sample spaces are called *discrete sample space*, while non–countably infinite sample spaces are called *continuous sample spaces*[1].

***Example* 7.** Consider an experiment in which a coin is tossed repeatedly until a head comes up for the first time.

In this experiment, head may come up on the first toss, or the second toss, or the third toss and so on. Thus for this experiment, we obtain an unending sequence of outcomes:

$$\mathcal{S} = \{H,\ TH,\ TTH,\ TTTH,\ \ldots\}.$$

This is an instance of countably infinite sample space.

***Example* 8.** An experiment is concerned with measuring the life spans of electric bulbs manufactured in a factory in a single day. In this case, we will get any real number between zero and a certain maximum value. Thus the sample space might be described by

$$\mathcal{S} = \{t \mid 0 \le t \le 4000\},$$

where t is measured in hours and the maximum life has been assumed to be 4000 hours.

♣ *Remark* 2. Generally, more than one sample space can describe the same random experiment, but frequently there will be only one, that will give the most information. For example the rolling of a fair die can be described by the sample spaces $\{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$ and {even, odd}, but the second sample space is not suitable if we required to find, for example, wether the number obtained will be divisible by 3.

2.2 Events

▷ **Definition 3 (Events).** An *event*, $\mathcal{A}$, is a subset of the sample space $\mathcal{S}$, i.e. a set $\mathcal{A}$ is an event *if and only if* $\mathcal{A} \subset \mathcal{S}$.

Thus the set $\{2, 4, 6\}$ is an event of the sample space $\{1, 2, 3, 4, 5, 6\}$ but $\{0, 1, 2\}$ is not.

Let $\mathcal{A}$ denote the event "a number less than 4 appears on the die". If actually a '1' appears on the die, then we will say that the event $\mathcal{A}$ has *occurred.* Thus, *whenever an outcome satisfies completely the conditions mentioned in the event, it is said to have occurred.* More specifically:

▷ **Definition 4** (*Occurrence of an event*)**.** The event $\mathcal{A}$ of the sample space $\mathcal{S}$ of an experiment is said to have occurred if the outcome (or the sample point) w of the experiment is such that $w \in \mathcal{A}$.

Alternatively, if $w \notin \mathcal{A}$, the event $\mathcal{A}$ has *not occurred.*

If an event $\mathcal{A}$ has only one sample point, it is called an *elementary* or *simple event.* If $\mathcal{A}$ consists of more than one point, it is called *compound event.* The sample space $\mathcal{S}$ is also a subset of itself, hence $\mathcal{S}$ is also an event. Moreover, any outcome of the experiment must be in $\mathcal{S}$. Hence, the sample space is called the *sure event.* The empty set $\varnothing$ is also a subset of the sample space, but the event $\varnothing$ can *never* occur. Hence, the empty subset of the sample space is called an *impossible event.* Finally, the *complement* of an event $\mathcal{A}$ with respect to the sample space $\mathcal{S}$, denoted by $\mathcal{A}'$, is the set of all elements of $\mathcal{S}$ which are not in $\mathcal{A}$:

$$\mathcal{A}' = \mathcal{S} - \mathcal{A} = \{w \mid w \in \mathcal{S} \text{ and } w \notin \mathcal{A}\}.$$

[1] We shall consider only finite sample spaces.

2.2.1 Algebra of Events

A sample space is a set (of all possible outcomes of an experiment), and the events are subsets of that set. Thus, sample space is a universal set for these events. Thus, the algebra of events is in effect the algebra of sets. Let $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be events of the sample space $\mathcal{S}$. Then,

1. $\mathcal{A} \cup \mathcal{B}$ is the event $\{\mathcal{A}$ or $\mathcal{B}$ or both$\}$.
2. $\mathcal{A} \cap \mathcal{B}$ is the event $\{$both $\mathcal{A}$ and $\mathcal{B}\}$.
3. $\mathcal{A}'$ is the event $\{$not $\mathcal{A}\}$.
4. $\mathcal{A} - \mathcal{B}$ is the event $\{\mathcal{A}$ but not $\mathcal{B}\}$.
5. $\mathcal{A} \cup \mathcal{B} = \mathcal{B} \cup \mathcal{A}$; $\quad \mathcal{A} \cap \mathcal{B} = \mathcal{B} \cap \mathcal{A}$.
6. $(\mathcal{A} \cup \mathcal{B})' = \mathcal{A}' \cap \mathcal{B}'$; $\quad (\mathcal{A} \cap \mathcal{B})' = \mathcal{A}' \cup \mathcal{B}'$.
7. $\mathcal{A} \cup (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cup \mathcal{C}$.
8. $\mathcal{A} \cap (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cap \mathcal{C}$.
9. $\mathcal{A} \cup (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cap (\mathcal{A} \cup \mathcal{C})$.
10. $\mathcal{A} \cap (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{C})$.

The above list is not exhaustive.

♣ *Remark* 3. If the sample space $\mathcal{S}$ has n sample points, the total number of events (i.e. total number of subsets of $\mathcal{S}$) is given by 2^n. This number includes the sure event $\mathcal{S}$ and the impossible event $\varnothing$.

2.2.2 Mutually Exclusive Events. Exhaustive Events

▷ **Definition 5** (**Mutually Exclusive Events**). Two events $\mathcal{A}$ and $\mathcal{B}$ are said to be mutually exclusive if the occurrence of one excludes the other. In the language of sets, $\mathcal{A}$ and $\mathcal{B}$ will be mutually exclusive if they are *disjoint* sets, i.e.

$$\mathcal{A} \cap \mathcal{B} = \varnothing. \tag{1}$$

More than two events are said to be mutually exclusive, if they are pairwise exclusive.

***Example* 9.** A card is drawn from a well shuffled deck of cards. Let the event $\mathcal{A} = \{$card is a heart$\}$ and $\mathcal{B} = \{$card is a diamond$\}$ are mutually exclusive because the same card cannot be a heart and a diamond at the same time. But the events $\mathcal{A} = \{$card is a heart$\}$ and $\mathcal{C} = \{$card is a king$\}$ are *not* mutually exclusive because there *is* a king of hearts.

♣ *Remark* 4. Elementary events of a sample space are always mutually exclusive.

♣ *Remark* 5. A standard deck of playing cards consists of 52 cards. The 52 cards are divided into *four* suits: *diamond* (♢), *heart* (♡), *club* (♣), and *spade* (♠). Each suit contains 13 cards: ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, queen, and king. The last three cards are known as *face cards* while the rest of them are *numbered*. Further, the heart and diamond suit are red in color and the club and the spade are black.

▷ **Definition 6** (**Exhaustive Events**). For a random experiment, a set of events is said to be exhaustive, if one of them must necessarily occur every time the experiment is performed. In the language of sets, the events $\mathcal{A}_1$, $\mathcal{A}_2$, ..., $\mathcal{A}_n$ of a sample space $\mathcal{S}$ describing the random experiment $\mathscr{E}$ are exhaustive if

$$\mathcal{A}_1 \cup \mathcal{A}_2 \cup \ldots \cup \mathcal{A}_n = \mathcal{S}. \tag{2}$$

Further, if

$$\mathcal{A}_i \cap \mathcal{A}_j = \varnothing \quad \forall \quad i \neq j,$$

then the events $\mathcal{A}_1$, $\mathcal{A}_2$, ..., $\mathcal{A}_n$ are called **mutually exclusive and exhaustive**.

3 Probability of an Event

We are now interested in determining the chances that a particular event will occur when we perform an experiment.

▷ **Definition 7** (**Equiprobable outcomes**). The outcomes of an experiment are said to be *equiprobable* or *equally likely* if there is no reason to believe that one is more likely to occur than the other.

For instance, if we roll a fair die, any of the number can occur on the top face. Since, the die is a homogenous cube, there is absolutely no reason to believe that a certain number, say 3, is more likely to show up than any other. So in this case, all outcomes are equiprobable. Similarly, if a card is drawn at random from a well shuffled deck of card, all outcomes have equal weights.

▷ **Definition 8** (**Probability**). If in an experiment with n equally likely (and exhaustive) outcomes, m outcomes are favorable to an event $\mathcal{A}$ (i.e. the event $\mathcal{A}$ can occur in m ways), then the probability[2] of occurrence of event $\mathcal{A}$, written as $\Pr(\mathcal{A})$, is given by

$$\Pr(\mathcal{A}) = \frac{\text{Number of outcomes favorable to } \mathcal{A}}{\text{Total number of possible outcomes}} = \frac{m}{n}. \tag{3}$$

If we denote the experiment as $\mathscr{E}$ and its sample space as $\mathcal{S}$, then obviously the total number of outcomes of $\mathscr{E}$ is the number of sample points, i.e. the number[3]of elements in $\mathcal{S} = |\mathcal{S}|$. A given event $\mathcal{A}$ will occur if the outcome of the experiment $w \in \mathcal{A}$ (obviously $w \in \mathcal{S}$). Thus the number of ways in which $\mathcal{A}$ can occur is nothing but $|\mathcal{A}|$. Hence

$$\Pr(\mathcal{A}) = \frac{|\mathcal{A}|}{|\mathcal{S}|}. \tag{4}$$

Also, by the above definition, the probability of an elementary event $\{w\}$:

$$\Pr(\{w\}) = \frac{|\{w\}|}{|\mathcal{S}|} = \frac{1}{|\mathcal{S}|}. \tag{5}$$

Thus, we can also write

$$\Pr(\mathcal{A}) = \sum_{a \in \mathcal{A}} \Pr(\{a\}). \tag{6}$$

♣ *Remark* 6. From the definition of probability, it follows:

(a) the probability of the sure event, $\Pr(\mathcal{S}) = |\mathcal{S}|/|\mathcal{S}| = 1$.

(b) the probability of the impossible event, $\Pr(\varnothing) = |\varnothing|/|\mathcal{S}| = 0/|\mathcal{S}| = 0$.

(c) Since $0 \le |\mathcal{A}| \le |\mathcal{S}|$, we must always have $0 \le \Pr(\mathcal{A}) \le 1$.

***Example* 10.** A card is drawn at random from a well–shuffled deck of 52 cards. Calculate the probability that the card will be (i) a diamond; (ii) an ace; (iii) not a ten.
*Solution:*In this case $|\mathcal{S}| = 52$.

(i) Let the event $\mathcal{A}$ ={card is diamond}. Thus $|\mathcal{A}|$ = no. of ways in 1 card can be selected out of 13 diamonds $= {}^{13}C_1 = 13$. Thus

$$\Pr(\mathcal{A}) = \frac{13}{52} = \frac{1}{4}.$$

[2]The definition given in this section is called the *classical definition of probability*. We shall be concerned with this definition only. The more general theory is based on the *axiomatic definition*.

[3]The number of elements in a set $\mathcal{A}$ will be denoted by $|\mathcal{A}|$ (read: "size of $\mathcal{A}$").

(ii) Let $\mathcal{B}$ ={card is an ace}. Since there are only 4 aces, 1 of them can be selected in 4 ways. Thus, $|\mathcal{B}| = 4$. Thus

$$\Pr(\mathcal{B}) = \frac{4}{52} = \frac{1}{13}.$$

(iii) Let $\mathcal{C}$ ={card is not ten}. Now, number of ways of selecting a ten is 4. So, the ways of *not* choosing any ace is the same as $52 - 4 = 48$. Thus,

$$\Pr(\mathcal{C}) = \frac{48}{52} = \frac{12}{13}.$$

***Example* 11.** Let there be a set $\mathcal{X}$. Find the probability of randomly selecting a subset of $\mathcal{X}$ having exactly k elements where $k \leq |\mathcal{X}|$.
Solution: The sample space $\mathcal{S}$ for this experiment is a set of all the subsets of $\mathcal{X}$ of all sizes that can be formed. Thus,

$$|\mathcal{S}| = 2^{|\mathcal{X}|}.$$

Let the event $\mathcal{A}$ ={subsets of $\mathcal{X}$ with size k}. Note that the event $\mathcal{A}$ consists of elements which themselves are sets (specifically, they are the subsets of $\mathcal{X}$ which consists of exactly k elements). Now

$$\begin{aligned} |\mathcal{A}| &= \text{no. of elements in } \mathcal{A} \\ &= \text{no. of subsets of } \mathcal{X} \text{ having exactly } k \text{ elements} \\ &= \text{no. of ways of selecting } k \text{ objects out of } |\mathcal{X}| \text{ distinct objects} \\ &= \binom{|\mathcal{X}|}{k}, \end{aligned}$$

where $\binom{n}{k}$ represents the binomial coefficient ${}^{n}\mathrm{C}_k$. Thus

$$\Pr(\mathcal{A}) = \binom{|\mathcal{X}|}{k} 2^{-|\mathcal{X}|}.$$

3.1 Some Useful Results

By using the algebra of events, some useful results can be used to solve the problems on probability. The results follow.

⊙ **Theorem 1 (Addition of Probabilities).** *For any two events $\mathcal{A}$ and $\mathcal{B}$ belonging to a sample space $\mathcal{S}$, the following is true:*

$$\Pr(\mathcal{A} \cup \mathcal{B}) = \Pr(\mathcal{A}) + \Pr(\mathcal{B}) - \Pr(\mathcal{A} \cap \mathcal{B}). \tag{7}$$

We shall not prove the theorem, but it can be intuitively clear by use of the *Venn diagram* shown in Fig. 5. It is obvious from the diagram that

$$|\mathcal{A} \cup \mathcal{B}| = |\mathcal{A}| + |\mathcal{B}| - |\mathcal{A} \cap \mathcal{B}|,$$

because while counting the elements of $\mathcal{A}$ we counted the shaded region once, then again while counting the elements of $\mathcal{B}$, and so it has been counted twice. To make things consistent, we must subtract the number present in the shaded region once. But the region is equal to the intersection of the two sets and so after dividing by the total number of elements in the sample space the result follows.

◇ **Corollary 1.** For mutually exclusive events $\mathcal{A}$ and $\mathcal{B}$ belonging to a sample space $\mathcal{S}$:

$$\Pr(\mathcal{A} \cup \mathcal{B}) = \Pr(\mathcal{A}) + \Pr(\mathcal{B}). \tag{8}$$

The above result follows quickly from (7) taking into account the facts that for mutually exclusive events $\mathcal{A}$ and $\mathcal{B}$: $\quad \mathcal{A} \cap \mathcal{B} = \varnothing \quad$ and $\quad \Pr(\varnothing) = 0$.

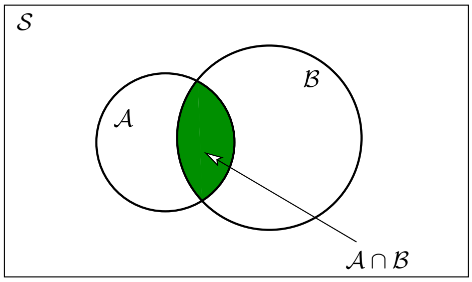

Figure 5: Venn diagram leading to the sum of probabilities.

⊙ **Theorem 2.** *If $\mathcal{A}'$ is the complement of $\mathcal{A}$, then*

$$\Pr(\mathcal{A}') = 1 - \Pr(\mathcal{A}). \tag{9}$$

The above theorem also follows from (8) noting the fact that $\mathcal{A}'$ and $\mathcal{A}$ are mutually exclusive and $\mathcal{A}\cup\mathcal{A}' = \mathcal{S}$. Also, $\Pr(\mathcal{S}) = 1$.

⊙ **Theorem 3.** *If $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n$, where $\mathcal{A}_i$'s are mutually exclusive events for $i = 1, 2, \ldots, n$, then*

$$\Pr(\mathcal{A}) = \Pr(\mathcal{A}_1) + \Pr(\mathcal{A}_2) + \cdots + \Pr(\mathcal{A}_n). \tag{10}$$

In particular, if $\mathcal{A} = \mathcal{S}$, then

$$\Pr(\mathcal{A}_1) + \Pr(\mathcal{A}_2) + \cdots + \Pr(\mathcal{A}_n) = \Pr(\mathcal{S}) = 1. \tag{11}$$

⊙ **Theorem 4.** *For any two events $\mathcal{A}$ and $\mathcal{B}$ of a sample space $\mathcal{S}$,*

$$\Pr(\mathcal{A}) = \Pr(\mathcal{A}\cap\mathcal{B}) + \Pr(\mathcal{A}\cap\mathcal{B}'). \tag{12}$$

The above can be easily seen if it is noted that

$$\mathcal{A} = \mathcal{A}\cup\varnothing = \mathcal{A}\cup(\mathcal{B}\cup\mathcal{B}') = (\mathcal{A}\cap\mathcal{B})\cup(\mathcal{A}\cap\mathcal{B}').$$

At the same time, the events $(\mathcal{A}\cap\mathcal{B})$ and $(\mathcal{A}\cap\mathcal{B}')$ are mutually exclusive because:

$$(\mathcal{A}\cap\mathcal{B})\cap(\mathcal{A}\cap\mathcal{B}') = \mathcal{A}\cap\mathcal{B}\cap\mathcal{A}\cap\mathcal{B}' = (\mathcal{A}\cap\mathcal{A})\cap(\mathcal{B}\cap\mathcal{B}') = \mathcal{A}\cap\varnothing = \varnothing.$$

Using (8) for the events $(\mathcal{A}\cap\mathcal{B})$ and $(\mathcal{A}\cap\mathcal{B}')$, the result (12) follows.

⊙ **Theorem 5.** *If $\mathcal{B}\subset\mathcal{A}$, then $\Pr(\mathcal{B}) \le \Pr(\mathcal{A})$ and $\Pr(\mathcal{A}-\mathcal{B}) = \Pr(\mathcal{A}) - \Pr(\mathcal{B})$.*

The above result follows from the fact that $\mathcal{A} = \mathcal{B}\cup(\mathcal{A}-\mathcal{B})$ where $\mathcal{B}$ and $(\mathcal{A}-\mathcal{B})$ are mutually exclusive. Thus $\Pr(\mathcal{A}) = \Pr(\mathcal{B}) + \Pr(\mathcal{A}-\mathcal{B})$ so that $\Pr(\mathcal{A}-\mathcal{B}) = \Pr(\mathcal{A}) - \Pr(\mathcal{B})$. Also since $\Pr(\mathcal{A}-\mathcal{B}) \ge 0$ so that $\Pr(\mathcal{B}) \le \Pr(\mathcal{A})$.

⊙ **Theorem 6.** *If an event $\mathcal{A}$ must result in the occurrence of one of the mutually exclusive events $\mathcal{A}_1$, $\mathcal{A}_2$, …, $\mathcal{A}_n$ then,*

$$\Pr(\mathcal{A}) = \Pr(\mathcal{A}\cap\mathcal{A}_1) + \Pr(\mathcal{A}\cap\mathcal{A}_2) + \cdots + \Pr(\mathcal{A}\cap\mathcal{A}_n). \tag{13}$$

Sometimes, the phrase "odds in favor" and "odds against" an event are used. These terms are closely related to probability.

Odds in favor. The *odds in favor of an event* is defined as ratio of the outcomes favorable to the event to that unfavorable to the same event. Thus

$$\begin{aligned}\text{Odds in favor of event } \mathcal{A} &= \frac{\text{no. of ways in which } \mathcal{A} \text{ can occur}}{\text{no. of ways in which } \mathcal{A} \text{ cannot occur}}\\ &= \frac{|\mathcal{A}|}{|\mathcal{S}| - |\mathcal{A}|}\\ &= \frac{\Pr(\mathcal{A})}{1 - \Pr(\mathcal{A})}.\end{aligned} \tag{14}$$

Odds against. The *odds again an event* is defined as ratio of the outcomes not favorable to the event to that favorable to the same event. Thus

$$\begin{aligned}\text{Odds against an event } \mathcal{A} &= \frac{\text{no. of ways in which } \mathcal{A} \text{ cannot occur}}{\text{no. of ways in which } \mathcal{A} \text{ can occur}}\\ &= \frac{|\mathcal{S}| - |\mathcal{A}|}{|\mathcal{A}|}\\ &= \frac{1 - \Pr(\mathcal{A})}{\Pr(\mathcal{A})}.\end{aligned} \tag{15}$$

***Example* 12.** A card is drawn at random from an ordinary deck of 52 playing cards. Find the probability that it is (i) a three of clubs or a six of diamonds; (ii) neither a four nor a club.
Solution: Let H, S, D and C denote, respectively, the suits of hearts, spades, diamonds and clubs and 1, 2, ... 13 for ace, two, three, ..., king. Then $3 \cap H$ means a three of hearts and $3 \cup H$ means a three or hearts.

(i) Required: $\Pr\big((3 \cap C) \cup (6 \cap D)\big)$. Since the event $(3 \cap C)$ and $(6 \cap D)$ are mutually exclusive, therefore

$$\Pr\big((3 \cap C) \cup (6 \cap D)\big) = \Pr(3 \cap C) + \Pr(6 \cap D) = \frac{1}{52} + \frac{1}{52} = \frac{1}{26}.$$

(ii) Required: $\Pr(4' \cap C')$. But $4' \cap C' = (4 \cup C)'$. Therefore,

$$\begin{aligned}\Pr(4' \cap C') &= \Pr\big((4 \cup C)'\big) = 1 - \Pr(4 \cup C)\\ &= 1 - \Big[\Pr(4) + \Pr(C) - \Pr(4 \cap C)\Big]\\ &= 1 - \left[\frac{1}{13} + \frac{1}{4} - \frac{1}{52}\right]\\ &= \frac{9}{13}\end{aligned}$$

3.2 Conditional Probability. Independent Events Multiplication of Probabilities

Suppose we have a well–shuffled deck of 52 cards and let the event $\mathcal{A} = \{\text{the top card is an ace}\}$. Then the probability of event $\mathcal{A}$ is given by

$$\Pr(\mathcal{A}) = \frac{4}{52} = \frac{1}{13}.$$

However, suppose we have noticed that the bottom card is an ace of hearts. What is now the probability that the top card is an ace?

We see that now there are only 51 cards left out (excluding the bottom card) of which there are only three aces. Therefore, the required probability that the top card is an ace given that the bottom card is an ace of hearts is just 3/51, which is not the same as before. In fact, this is a bit smaller than the earlier. This is to be expected because now that we know that the bottom is an ace, the probability of the top being an ace should be lower.

To distinguish this probability from the original probability of the event $\Pr(\mathcal{A})$, we denote it by $\Pr(\mathcal{A}|\mathcal{B})$ and call it the *conditional probability* of the event $\mathcal{A}$ given that the event $\mathcal{B} = \{$the bottom card is an ace of hearts$\}$.

We see that $\Pr(\mathcal{A})$ and $\Pr(\mathcal{A}|\mathcal{B})$ are not the same. Now the question arises how to calculate the conditional probability of events.

To this end, look at the Venn diagram in Fig. 5. Let it be known that event $\mathcal{B}$ has occurred and we sought to find $\Pr(\mathcal{A}|\mathcal{B})$. Since event $\mathcal{B}$ has already occurred, the effective sample space for event $\mathcal{A}$ in the light of event $\mathcal{B}$ has reduced to $\mathcal{B}$. Thus,

$$\Pr(\mathcal{A}|\mathcal{B}) = \frac{|\mathcal{A} \cap \mathcal{B}|}{|\mathcal{B}|}.$$

Dividing the numerator and denominator by $|\mathcal{S}|$, we obtain the definition of conditional probability as

$$\Pr(\mathcal{A}|\mathcal{B}) \equiv \frac{\Pr(\mathcal{A} \cap \mathcal{B})}{\Pr(\mathcal{B})}. \qquad \Big(\Pr(\mathcal{B}) \neq 0\Big) \tag{16}$$

Alternatively, we obtain

$$\Pr(\mathcal{A} \cap \mathcal{B}) = \Pr(\mathcal{B})\ \Pr(\mathcal{A}|\mathcal{B}) \tag{17}$$

In words, (17) says that *the probability that both the events, $\mathcal{A}$ and $\mathcal{B}$, occur is equal to the probability that one of them, say $\mathcal{A}$, occurs times the conditional probability of occurrence of the other one $\mathcal{B}$, given $\mathcal{A}$.*

▷ **Definition 9 (Independent Events).** If the probability of occurrence of an event $\mathcal{A}$ is not affected by the occurrence or non–occurrence of event $\mathcal{B}$ associated with the same sample space, then the events $\mathcal{A}$ and $\mathcal{B}$ are said to be *independent.*

We have

$$\text{for independent events} \qquad \Pr(\mathcal{A}|\mathcal{B}) = \Pr(\mathcal{A}). \tag{18}$$

And hence, by (17) and (18), for independent events, we have the rule for **multiplication of probabilities:**

$$\Pr(\mathcal{A} \cap \mathcal{B}) = \Pr(\mathcal{A})\,\Pr(\mathcal{B}). \tag{19}$$

♣ *Remark* 7.

1. In expression (16), it is assumed that $\Pr(\mathcal{B}) \neq 0$. On the other hand, if $\Pr(\mathcal{B}) = 0$, then the right hand side of this expression is not defined. Besides, if $\Pr(\mathcal{B}) = 0$, then $\mathcal{B}$ is an impossible event, so that the assumption that it has occurred is meaningless.
2. For the discussion resulting in conditional probability, we have assumed that each elementary event of $\mathcal{S}$ is equally likely. However, even if this assumption is not valid, the definition of conditional probability remains valid.
3. There should be no confusion regarding *exclusive* and *independent* events. Exclusiveness is a set–theoretic notion (whether we are talking of probability or not), whereas independence is related to the notion of occurrence of events and hence directly related to the concept of probability.

***Example* 13.** A box contains 5 red marbles and 4 white marbles. Two marbles are drawn successively from the box without replacement, and it is noted that the second one is white. What is the probability that the first is also white?

Solution: **Method 1.** If $\mathcal{W}_1$ and $\mathcal{W}_2$ be the events $\{$white on first draw$\}$ and $\{$white on 2nd draw$\}$ respectively, we are looking for $\Pr(\mathcal{W}_1|\mathcal{W}_2)$. This is given by

$$\Pr(\mathcal{W}_1|\mathcal{W}_2) = \frac{\Pr(\mathcal{W}_1 \cap \mathcal{W}_2)}{\Pr(\mathcal{W}_2)} = \frac{(4/9)(3/8)}{4/9} = \frac{3}{8}.$$

Method 2. Since the second is known to be white, there are only 3 ways out of remaining 8 in which the first can be white, so that the probability is 3/8.

***Example* 14.** A box contains 6 red, 4 white and 5 black balls. Three balls are successively drawn from it. Find the probability that they are drawn in order red, white and black, if each ball is (i) replaced, (ii) not replaced.
Solution: Let $\mathcal{R}_1 = \{\text{red on first draw}\}$, $\mathcal{W}_2 = \{\text{white on second draw}\}$, $\mathcal{B}_3 = \{\text{black on third draw}\}$. We require $\Pr(\mathcal{R}_1 \cap \mathcal{W}_2 \cap \mathcal{B}_3)$.
(i) If each ball is replaced, then the events are independent and

$$\begin{aligned}\Pr(\mathcal{R}_1 \cap \mathcal{W}_2 \cap \mathcal{B}_3) &= \Pr(\mathcal{R}_1)\Pr(\mathcal{W}_2|\mathcal{R}_1)\Pr(\mathcal{B}_3|\mathcal{R}_1 \cap \mathcal{W}_2)\\ &= \Pr(\mathcal{R}_1)\Pr(\mathcal{W}_2)\Pr(\mathcal{B}_3)\\ &= \left(\frac{6}{6+4+5}\right)\left(\frac{4}{6+4+5}\right)\left(\frac{5}{6+4+5}\right) = \frac{8}{225}\end{aligned}$$

(ii) If each of the balls is not replaced, then the events are dependent and

$$\begin{aligned}\Pr(\mathcal{R}_1 \cap \mathcal{W}_2 \cap \mathcal{B}_3) &= \Pr(\mathcal{R}_1)\Pr(\mathcal{W}_2|\mathcal{R}_1)\Pr(\mathcal{B}_3|\mathcal{R}_1 \cap \mathcal{W}_2)\\ &= \left(\frac{6}{6+4+5}\right)\left(\frac{4}{5+4+5}\right)\left(\frac{5}{5+3+5}\right) = \frac{4}{91}.\end{aligned}$$

The following theorems regarding conditional probability are important.

⊙ **Theorem 7.** *For three events $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$ of a sample space, we have*

$$\Pr(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3) = \Pr(\mathcal{A}_1)\Pr(\mathcal{A}_2|\mathcal{A}_1)\Pr(\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2). \tag{20}$$

The result may be generalized for n events.

▷ **Definition 10.** Three events $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$ of a sample space are said to be *independent*, if they are pairwise independent:

$$\Pr(\mathcal{A}_i \cap \mathcal{A}_j) = \Pr(\mathcal{A}_i)\Pr(\mathcal{A}_j) \qquad i \neq j \qquad \text{where} \quad i,\, j = 1, 2, 3 \tag{21}$$

$$\text{and} \qquad \Pr\left(\bigcap_{i=1}^{3} \mathcal{A}_i\right) = \prod_{i=1}^{3} \Pr(\mathcal{A}_i) \tag{22}$$

Note that neither (21) nor (22) is by itself sufficient. Independence of more than three events is easily defined.

Finally, related to the conditional probability is the *law of total probability* which is stated in form of a theorem:

⊙ **Theorem 8 (Law of Total Probability).** *If an event $\mathcal{A}$ must occur with one of the mutually exclusive and exhaustive events $\mathcal{A}_1$, $\mathcal{A}_2$, ..., $\mathcal{A}_n$, then*

$$\Pr(\mathcal{A}) = \Pr(\mathcal{A}_1)\Pr(\mathcal{A}|\mathcal{A}_1) + \Pr(\mathcal{A}_2)\Pr(\mathcal{A}|\mathcal{A}_2) + \ldots + \Pr(\mathcal{A}_n)\Pr(\mathcal{A}|\mathcal{A}_n) \tag{23}$$

***Example* 15.** A person has undertaken a construction job. The probabilities are 0.65 that there will be a strike, 0.80 that the construction job will be completed on time if there is no strike, and 0.32 that the construction job will be completed on time if there *is* a strike. Determine the probability that the construction job will be finished on time.
Solution: Let $\mathcal{A} = \{\text{the construction job will be finished on time}\}$ and $\mathcal{B} = \{\text{there will be a strike}\}$. We require $\Pr(\mathcal{A})$. We have, in our notations, the following pieces of information:

$$\Pr(\mathcal{B}) = 0.65, \qquad \Pr(\mathcal{A}|\mathcal{B}') = 0.80, \qquad \Pr(\mathcal{A}|\mathcal{B}) = 0.32.$$

Here, $\mathcal{A}$ is an event which must occur with either $\mathcal{B}$ or $\mathcal{B}'$ which are mutually exclusive and exhaustive. Thus by the theorem on total probability, we have

$$\begin{aligned}\Pr(\mathcal{A}) &= \Pr(\mathcal{B})\Pr(\mathcal{A}|\mathcal{B}) + \Pr(\mathcal{B}')\Pr(\mathcal{A}|\mathcal{B}')\\ &= \Pr(\mathcal{B})\Pr(\mathcal{A}|\mathcal{B}) + \big(1 - \Pr(\mathcal{B})\big)\Pr(\mathcal{A}|\mathcal{B}')\\ &= 0.65 \times 0.32 + (1 - 0.65) \times 0.8\\ &= 0.488\end{aligned}$$

3.3 Bayes' Theorem or Rule

⊙ **Theorem 9 (Bayes' theorem).** *Let* $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \ldots, \mathcal{A}_n$ *be* mutually exclusive and exhaustive *events. If* $\mathcal{A}$ *be* any *event of the same sample space, then*

$$\Pr(\mathcal{A}_k|\mathcal{A}) = \frac{\Pr(\mathcal{A}_k)\ \Pr(\mathcal{A}|\mathcal{A}_k)}{\sum_{j=1}^{n}\Pr(\mathcal{A}_j)\ \Pr(\mathcal{A}|\mathcal{A}_j)}. \tag{24}$$

Proof. Since $\mathcal{A}_i$'s are mutually exclusive and exhaustive events, we have

$$\mathcal{A}_i \cap \mathcal{A}_j = \varnothing \quad \forall \quad i \neq j \qquad \text{and} \qquad \bigcup_{i=1}^{n} \mathcal{A}_i = \mathcal{S}.$$

Now, if $\mathcal{A}$ be any event of the same sample space, we have

$$\mathcal{A} = \mathcal{A}\cap\mathcal{S} = \mathcal{A}\cap\big(\mathcal{A}_1\cup\mathcal{A}_2\cup\cdots\cup\mathcal{A}_n\big) = (\mathcal{A}\cap\mathcal{A}_1)\cup(\mathcal{A}\cap\mathcal{A}_2)\cup\cdots\cup(\mathcal{A}\cap\mathcal{A}_n).$$

Also, each pair $(\mathcal{A}\cap\mathcal{A}_i)$ and $(\mathcal{A}\cap\mathcal{A}_j)$ are mutually exclusive, for we have

$$\begin{aligned}(\mathcal{A}\cap\mathcal{A}_i)\cap(\mathcal{A}\cap\mathcal{A}_j) &= \mathcal{A}\cap\mathcal{A}_i\cap\mathcal{A}\cap\mathcal{A}_j \\ &= \mathcal{A}\cap(\mathcal{A}_i\cap\mathcal{A})\cap\mathcal{A}_j \\ &= \mathcal{A}\cap(\mathcal{A}\cap\mathcal{A}_i)\cap\mathcal{A}_j \\ &= (\mathcal{A}\cap\mathcal{A})\cap(\mathcal{A}_i\cap\mathcal{A}_j) \\ &= \mathcal{A}\cap\varnothing = \varnothing.\end{aligned}$$

$$\begin{aligned}\therefore\quad \Pr(\mathcal{A}) &= \Pr(\mathcal{A}\cap\mathcal{A}_1)+\Pr(\mathcal{A}\cap\mathcal{A}_2)+\cdots+\Pr(\mathcal{A}\cap\mathcal{A}_n) = \sum_{j=1}^{n}\Pr(\mathcal{A}\cap\mathcal{A}_j) \\ &= \sum_{j=1}^{n}\Pr(\mathcal{A}_j)\ \Pr(\mathcal{A}|\mathcal{A}_j)\,.\end{aligned}$$

Finally,

$$\Pr(\mathcal{A}_k|\mathcal{A}) = \frac{\Pr(\mathcal{A}_k\cap\mathcal{A})}{\Pr(\mathcal{A})} = \frac{\Pr(\mathcal{A}_k)\ \Pr(\mathcal{A}|\mathcal{A}_k)}{\sum_{j=1}^{n}\Pr(\mathcal{A}_j)\ \Pr(\mathcal{A}|\mathcal{A}_j)}\ .$$

□

This theorem enables us to find the probabilities of the various events $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \ldots, \mathcal{A}_n$ that can *cause* event $\mathcal{A}$ to occur. For this reason, Bayes' theorem is also referred to as a *theorem on the probability of causes.*

4 Random Variables. Probability Distributions

4.1 Random Variables

Random variables are neither random nor variables. They are functions defined on the sample space of a random experiment that map each point of the sample space to the set of real numbers. They are usually denoted by capital letters at the tail of alphabets, viz. X, Y, etc.

▷ **Definition 11 (Finite Probability Space).** A finite probability space is a finite set $\mathcal{S}$ together with a function $\Pr : \mathcal{S} \longrightarrow \mathbb{R}$ such that

$$0 \le \Pr(s) \le 1 \quad \forall \quad s\in\mathcal{S} \qquad \text{and} \qquad \sum_{s\in\mathcal{S}}\Pr(s) = 1.$$

We shall denote probability spaces as $\big(\mathcal{S},\,\Pr\big)$.

It is obvious that the elements, $s \in \mathcal{S}$ are elementary events of the sample space $\mathcal{S}$. The function Pr is called the probability function. We note that our earlier definition of probability is a special case of the more general definition given in this section. More specifically, if the function Pr associate the same number $1/|\mathcal{S}|$ to each sample point, it coincides with the classical definition.
The definition of random variable follows.

▷ **Definition 12** (**Random Variable**)**.** Given a probability space $\big(\mathcal{S},\,\Pr\big)$, a *random variable* on this space is a function $X : \mathcal{S} \longrightarrow \mathbb{R}$.

In general, random variable has some specified physical, geometrical, or other significance.

***Example* 16.** Suppose that a coin is tossed twice so that the sample space is $\mathcal{S} = \{HH,\,HT,\,TH,\,TT\}$. Let X represent the number of heads that can come. With each sample point we can associate a number for X as shown in Table 1. Thus, for example, in the case HH (i.e. two heads), $X = 2$ while for TH (one head), $X = 1$. It follows that X is a random variable.

Table 1: The number of heads in two tosses of a fair coin.

Sample point	HH	HT	TH	TT
X	2	1	1	0

Table 2: The probability distribution for Example 16.

x	0	1	2
$f(x)$	1/4	1/2	1/4

It should be noted that many random variables can be defined on the same sample space. For example, in Example 16, we could also have defined X as representing the square of the number of heads or the number of heads minus the number of tails.

A random variable that takes on a finite or countably infinite number of values is called a *discrete random variables* while one which takes on a non countable infinite number of values is called *continuous random variable.*

4.2 Discrete Probability Distributions

Let X be a discrete random variable, and suppose the possible values that it can take are denoted by x_1, x_2, x_3, ..., arranged in some order. Suppose also that these values are taken with probabilities given by

$$\Pr(X = x_k) = f(x_k) \qquad k = 1, 2, 3, \ldots \tag{25}$$

It is then convenient to introduce the *probability function*, also referred to as the *probability distribution*, given by

$$\Pr(X = x) = f(x). \tag{26}$$

For $x = x_k$, this reduced to (25) while for other values of x, $f(x) = 0$.

In general, $f(x)$ is a probability distribution if

1. $f(x) \geq 0$, and
2. $\sum_x f(x) = 1$, where the sum is taken over all possible values of x.

***Example* 17.** Find the probability distribution corresponding to the random variable X in Example 16. Assuming that the coin is fair, we have

$$\Pr(HH) = \frac{1}{4}, \quad \Pr(HT) = \frac{1}{4}, \quad \Pr(TH) = \frac{1}{4}, \quad \Pr(TT) = \frac{1}{4}.$$

Then

$$\begin{aligned} \Pr(X = 0) &= \Pr(TT) = \frac{1}{4} \\ \Pr(X = 1) &= \Pr(HT \cup TH) = \Pr(HT) + \Pr(TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \\ \Pr(X = 2) &= \Pr(HH) = \frac{1}{4} \end{aligned}$$

The probability distribution is thus given by Table 2

***Example* 18.** Suppose that a pair of fair dice are rolled once and let the random variable X denote the sum of the points. Obtain the probability distribution for X.
The sample points for this experiment is shown in Fig. 2. The random variable is the sum for of the coordinates for each point. Thus for $(3, 2)$ we have $X = 5$. Using the fact that all 36 sample points are equiprobable, so that each sample point has equal probability of 1/36, we obtain the Table 3. For example, corresponding to $X = 5$, we have the sample points (1,4), (2,3), (3,2) and (4,1), so that the associated probability is 4/36.

Table 3: Probability distribution for Example 18.

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

4.3 Cumulative Distribution Functions

▷ **Definition 13.** The *cumulative distribution function* (CDF), or briefly the *distribution function*, for a random variable X is defined by

$$F(x) = \Pr(X \leq x) \qquad \forall \quad x \in \mathbb{R}. \tag{27}$$

For a discrete random variable X, the CDF can be obtained from its probability distribution by noting that, for all x in the interval $(-\infty, \infty)$

$$F(x) = \Pr(X \leq x) = \sum_{u \leq x} f(u), \tag{28}$$

where the sum is taken over all values u taken on by X for which $u \leq x$.

Further, if X takes on only a finite number of values $x_1, x_2, \ldots, x_n$, its CDF is given by

$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ f(x_1) & x_1 \leq x < x_2 \\ f(x_1) + f(x_2) & x_2 \leq x < x_3 \\ \quad\vdots & \quad\vdots \\ f(x_1) + f(x_2) + \cdots + f(x_n) & x_n \leq x < \infty \end{cases} \tag{29}$$

From its definition, the following properties of CDFs can be easily deduced:

1. $F(x)$ is non–decreasing, i.e. for all $x \le y$, $F(x) \le F(y)$.

2. $\lim_{x\to-\infty} F(x) = 0$; $\lim_{x\to\infty} F(x) = 1$.

3. $F(x)$ is continuous from the right at all points, i.e. $\lim_{x\to x_0+} F(x) = F(x_0)$ for all $x_0 \in \mathbb{R}$.

***Example* 19.** (a) Find the CDF for the random variable X of Example 17. (b) Obtain its graph.
Solution: (a) The CDF is as follows:

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

(b) The graph of $F(x)$ is shown in Fig. 6.

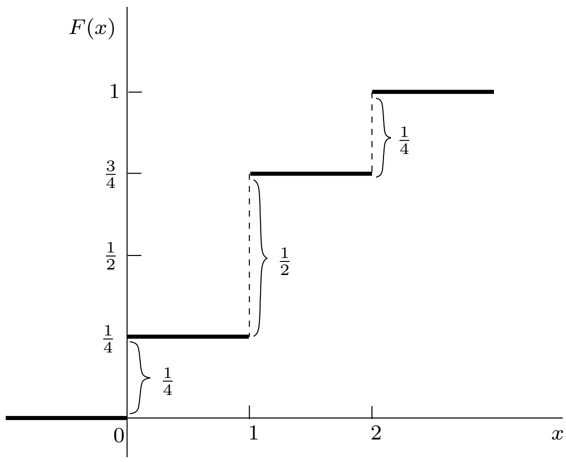

Figure 6: Graph of CDF in Example 19.

The following things about Fig. 6, *which are true in general*, should be noted:

1. The magnitudes of the jumps at 0, 1, 2 are $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ respectively which are precisely the probabilities in Table 2. This fact enables one to obtain the probability distribution from the CDF.

2. The graph of a discrete CDF is a *step* function.

3. As we proceed from left to right (i.e. *climb up the steps*), the CDF either remains the same or increases, taking on values from 0 to 1. Thus a CDF is *monotonically increasing function.*

From the above, it is clear that the probability distribution of a discrete random variable can be obtained from its CDF by noting that

$$f(x) = F(x) - \lim_{u\to x-} F(u). \tag{30}$$