

**MODULE 1**  
**PROBABILITY**  
**LECTURES 1-6**

**Topics**

**1.1 INTRODUCTION**

*1.1.1 Classical Method*

*1.1.2 Relative Frequency Method*

**1.2 AXIOMATIC APPROACH TO PROBABILITY AND  
PROPERTIES OF PROBABILITY MEASURE**

*1.2.1 Inclusion-Exclusion Formula*

*1.2.1.1 Boole's Inequality*

*1.2.1.2 Bonferroni's Inequality*

*1.2.2 Equally Likely Probability Models*

**1.3 CONDITIONAL PROBABILITY AND  
INDEPENDENCE OF EVENTS**

*1.3.1 Theorem of Total Probability*

*1.3.2 Bayes' Theorem*

**1.4 CONTINUITY OF PROBABILITY MEASURES**

**MODULE 1**

**PROBABILITY**

**LECTURE 1**

**Topics**

**1.1 INTRODUCTION**

*1.1.1 Classical Method*

*1.1.2 Relative Frequency Method*

## 1.1 INTRODUCTION

In our daily life we come across many processes whose nature cannot be predicted in advance. Such processes are referred to as *random processes*. The only way to derive information about random processes is to conduct experiments. Each such experiment results in an outcome which cannot be predicted beforehand. In fact even if the experiment is repeated under identical conditions, due to presence of factors which are beyond control, outcomes of the experiment may vary from trial to trial. However we may know in advance that each outcome of the experiment will result in one of the several given possibilities. For example, in the cast of a die under a fixed environment the outcome (number of dots on the upper face of the die) cannot be predicted in advance and it varies from trial to trial. However we know in advance that the outcome has to be among one of the numbers  $1, 2, \dots, 6$ . *Probability theory* deals with the modeling and study of random processes. The field of *Statistics* is closely related to probability theory and it deals with drawing inferences from the data pertaining to random processes.

### Definition 1.1

- (i) A *random experiment* is an experiment in which:
  - (a) the set of all possible outcomes of the experiment is known in advance;
  - (b) the outcome of a particular performance (trial) of the experiment cannot be predicted in advance;
  - (c) the experiment can be repeated under identical conditions.
- (ii) The collection of all possible outcomes of a random experiment is called the *sample space*. A sample space will usually be denoted by  $\Omega$ . ■

### Example 1.1

- (i) In the random experiment of casting a die one may take the sample space as  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , where  $i \in \Omega$  indicates that the experiment results in  $i$  ( $i = 1, \dots, 6$ ) dots on the upper face of die.
- (ii) In the random experiment of simultaneously flipping a coin and casting a die one may take the sample space as

$$\Omega = \{H, T\} \times \{1, 2, \dots, 6\} = \{(r, i) : r \in \{H, T\}, i \in \{1, 2, \dots, 6\}\}$$

where  $(H, i)$  ( $(T, i)$ ) indicates that the flip of the coin resulted in head (tail) on the upper face and the cast of the die resulted in  $i$  ( $i = 1, 2, \dots, 6$ ) dots on the upper face.

- (iii) Consider an experiment where a coin is tossed repeatedly until a head is observed. In this case the sample space may be taken as  $\Omega = \{1, 2, \dots\}$  (or

- $\Omega = \{T, TH, TTH, \dots\}$ ), where  $i \in \Omega$  (or  $TT \dots TH \in \Omega$  with  $(i - 1)$  Ts and one H) indicates that the experiment terminates on the  $i$ -th trial with first  $i - 1$  trials resulting in tails on the upper face and the  $i$ -th trial resulting in the head on the upper face.
- (iv) In the random experiment of measuring lifetimes (in hours) of a particular brand of batteries manufactured by a company one may take  $\Omega = [0, 70,000]$ , where we have assumed that no battery lasts for more than 70,000 hours. ■

### Definition 1.2

- (i) Let  $\Omega$  be the sample space of a random experiment and let  $E \subseteq \Omega$ . If the outcome of the random experiment is a member of the set  $E$  we say that the *event*  $E$  has occurred.
- (ii) Two events  $E_1$  and  $E_2$  are said to be *mutually exclusive* if they cannot occur simultaneously, i.e., if  $E_1 \cap E_2 = \phi$ , the empty set. ■

In a random experiment some events may be more likely to occur than the others. For example, in the cast of a fair die (a die that is not biased towards any particular outcome), the occurrence of an odd number of dots on the upper face is more likely than the occurrence of 2 or 4 dots on the upper face. Thus it may be desirable to quantify the likelihoods of occurrences of various events. *Probability* of an event is a numerical measure of chance with which that event occurs. To assign probabilities to various events associated with a random experiment one may assign a real number  $P(E) \in [0, 1]$  to each event  $E$  with the interpretation that there is a  $(100 \times P(E))\%$  chance that the event  $E$  will occur and a  $(100 \times (1 - P(E)))\%$  chance that the event  $E$  will not occur. For example if the probability of an event is 0.25 it would mean that there is a 25% chance that the event will occur and that there is a 75% chance that the event will not occur. Note that, for any such assignment of possibilities to be meaningful, one must have  $P(\Omega) = 1$ . Now we will discuss two methods of assigning probabilities.

#### 1.1.1 Classical Method

This method of assigning probabilities is used for random experiments which result in a finite number of equally likely outcomes. Let  $\Omega = \{\omega_1, \dots, \omega_n\}$  be a finite sample space with  $n$  ( $\in \mathbb{N}$ ) possible outcomes; here  $\mathbb{N}$  denotes the set of natural numbers. For  $E \subseteq \Omega$ , let  $|E|$  denote the number of elements in  $E$ . An outcome  $\omega \in \Omega$  is said to be favorable to an event  $E$  if  $\omega \in E$ . In the classical method of assigning probabilities, the probability of an event  $E$  is given by

$$P(E) = \frac{\text{number of outcomes favorable to } E}{\text{total number of outcomes}} = \frac{|E|}{|\Omega|} = \frac{|E|}{n}.$$

Note that probabilities assigned through classical method satisfy the following properties of intuitive appeal:

- (i) For any event  $E$ ,  $P(E) \geq 0$ ;
- (ii) For mutually exclusive events  $E_1, E_2, \dots, E_n$  (i.e.,  $E_i \cap E_j = \phi$ , whenever  $i, j \in \{1, \dots, n\}, i \neq j$ )

$$P\left(\bigcup_{i=1}^n E_i\right) = \frac{|\bigcup_{i=1}^n E_i|}{n} = \frac{\sum_{i=1}^n |E_i|}{n} = \sum_{i=1}^n \frac{|E_i|}{n} = \sum_{i=1}^n P(E_i);$$

- (iii)  $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$ .

### Example 1.2

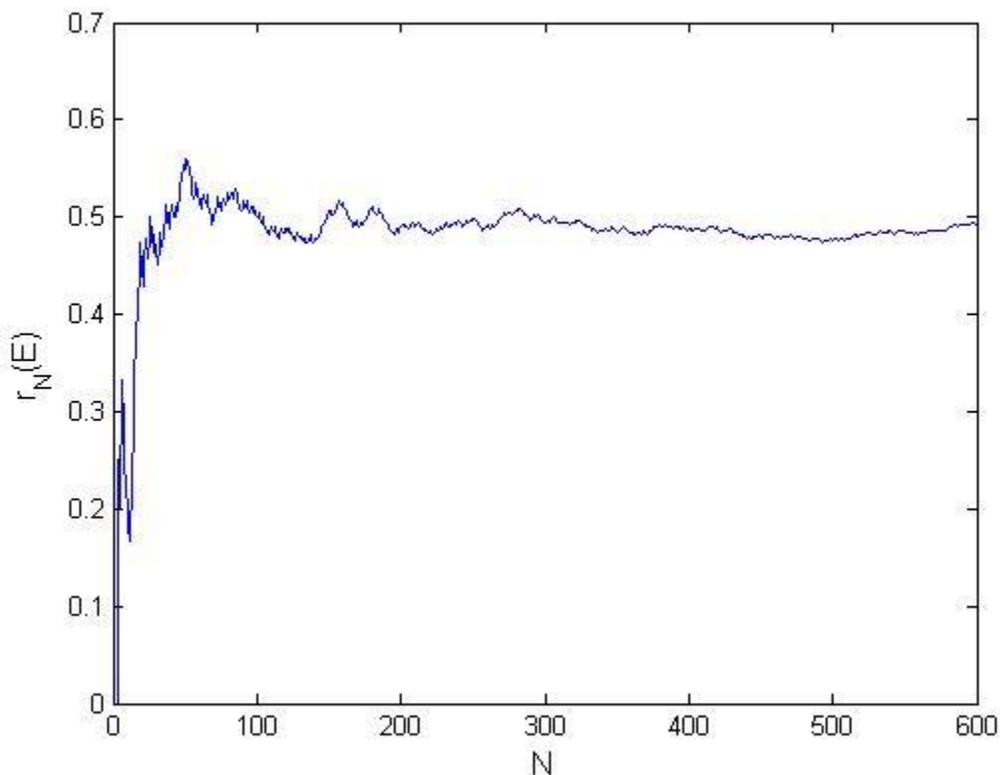
Suppose that in a classroom we have 25 students (with registration numbers 1, 2, ..., 25) born in the same year having 365 days. Suppose that we want to find the probability of the event  $E$  that they all are born on different days of the year. Here an outcome consists of a sequence of 25 birthdays. Suppose that all such sequences are equally likely. Then  $|\Omega| = 365^{25}$ ,  $|E| = 365 \times 364 \times \dots \times 341 = 365 P_{25}$  and  $P(E) = \frac{|E|}{|\Omega|} = \frac{365 P_{25}}{365^{25}}$ . ■

The classical method of assigning probabilities has a limited applicability as it can be used only for random experiments which result in a finite number of equally likely outcomes. ■

#### 1.1.2 Relative Frequency Method

Suppose that we have independent repetitions of a random experiment (here independent repetitions means that the outcome of one trial is not affected by the outcome of another trial) under identical conditions. Let  $f_N(E)$  denote the number of times an event  $E$  occurs (also called the frequency of event  $E$  in  $N$  trials) in the first  $N$  trials and let  $r_N(E) = f_N(E)/N$  denote the corresponding relative frequency. Using advanced probabilistic arguments (e.g., using Weak Law of Large Numbers to be discussed in Module 7) it can be shown that, under mild conditions, the relative frequencies stabilize (in certain sense) as  $N$  gets large (i.e., for any event  $E$ ,  $\lim_{N \rightarrow \infty} r_N(E)$  exists in certain sense). In the relative frequency method of assigning probabilities the probability of an event  $E$  is given by

$$P(E) = \lim_{N \rightarrow \infty} r_N(E) = \lim_{N \rightarrow \infty} \frac{f_N(E)}{N}.$$



**Figure 1.1.** Plot of relative frequencies ( $r_N(E)$ ) of number of heads against number of trials ( $N$ ) in the random experiment of tossing a fair coin (with probability of head in each trial as 0.5).

In practice, to assign probability to an event  $E$ , the experiment is repeated a large (but fixed) number of times (say  $N$  times) and the approximation  $P(E) \approx r_N(E)$  is used for assigning probability to event  $E$ . Note that probabilities assigned through relative frequency method also satisfy the following properties of intuitive appeal:

- (i) for any event  $E$ ,  $P(E) \geq 0$ ;
- (ii) for mutually exclusive events  $E_1, E_2, \dots, E_n$

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i);$$

- (iii)  $P(\Omega) = 1$ .

Although the relative frequency method seems to have more applicability than the classical method it too has limitations. A major problem with the relative frequency method is that it is imprecise as it is based on an approximation ( $P(E) \approx r_N(E)$ ).

Another difficulty with relative frequency method is that it assumes that the experiment can be repeated a large number of times. This may not be always possible due to budgetary and other constraints (e.g., in predicting the success of a new space technology it may not be possible to repeat the experiment a large number of times due to high costs involved).

The following definitions will be useful in future discussions.

**Definition 1.3**

- (i) A set  $E$  is said to be *finite* if either  $E = \phi$  (the empty set) or if there exists a one-one and onto function  $f: \{1, 2, \dots, n\} \rightarrow E$  (or  $f: E \rightarrow \{1, 2, \dots, n\}$ ) for some natural number  $n$ ;
- (ii) A set is said to be *infinite* if it is not finite;
- (iii) A set  $E$  is said to be *countable* if either  $E = \phi$  or if there is an onto function  $f: \mathbb{N} \rightarrow E$ , where  $\mathbb{N}$  denotes the set of natural numbers;
- (iv) A set is said to be *countably infinite* if it is countable and infinite;
- (v) A set is said to be *uncountable* if it is not countable;
- (vi) A set  $E$  is said to be *continuum* if there is a one-one and onto function  $f: \mathbb{R} \rightarrow E$  (or  $f: E \rightarrow \mathbb{R}$ ), where  $\mathbb{R}$  denotes the set of real numbers. ■

The following proposition, whose proof (s) can be found in any standard textbook on set theory, provides some of the properties of finite, countable and uncountable sets.

**Proposition 1.1**

- (i) Any finite set is countable;
- (ii) If  $A$  is a countable and  $B \subseteq A$  then  $B$  is countable;
- (iii) Any uncountable set is an infinite set;
- (iv) If  $A$  is an infinite set and  $A \subseteq B$  then  $B$  is infinite;
- (v) If  $A$  is an uncountable set and  $A \subseteq B$  then  $B$  is uncountable;
- (vi) If  $E$  is a finite set and  $F$  is a set such that there exists a one-one and onto function  $f: E \rightarrow F$  (or  $f: F \rightarrow E$ ) then  $F$  is finite;
- (vii) If  $E$  is a countably infinite (continuum) set and  $F$  is a set such that there exists a one-one and onto function  $f: E \rightarrow F$  (or  $f: F \rightarrow E$ ) then  $F$  is countably infinite (continuum);
- (viii) A set  $E$  is countable if and only if either  $E = \phi$  or there exists a one-one and onto map  $f: E \rightarrow \mathbb{N}_0$ , for some  $\mathbb{N}_0 \subseteq \mathbb{N}$ ;
- (ix) A set  $E$  is countable if, and only if, either  $E$  is finite or there exists a one-one map  $f: \mathbb{N} \rightarrow E$ ;

- (x) A set  $E$  is countable if, and only if, either  $E = \phi$  or there exists a one-one map  $f: E \rightarrow \mathbb{N}$ ;
- (xi) A non-empty countable set  $E$  can be either written as  $E = \{\omega_1, \omega_2, \dots, \omega_n\}$ , for some  $n \in \mathbb{N}$ , or as  $E = \{\omega_1, \omega_2, \dots\}$ ;
- (xii) Unit interval  $(0,1)$  is uncountable. Hence any interval  $(a, b)$ , where  $-\infty < a < b < \infty$ , is uncountable;
- (xiii)  $\mathbb{N} \times \mathbb{N}$  is countable;
- (xiv) Let  $\Lambda$  be a countable set and let  $\{A_\alpha: \alpha \in \Lambda\}$  be a (countable) collection of countable sets. Then  $\bigcup_{\alpha \in \Lambda} A_\alpha$  is countable. In other words, countable union of countable sets is countable;
- (xv) Any continuum set is uncountable. ■

### Example 1.3

- (i) Define  $f: \mathbb{N} \rightarrow \mathbb{N}$  by  $f(n) = n, n \in \mathbb{N}$ . Clearly  $f: \mathbb{N} \rightarrow \mathbb{N}$  is one-one and onto. Thus  $\mathbb{N}$  is countable. Also it can be easily seen (using the contradiction method) that  $\mathbb{N}$  is infinite. Thus  $\mathbb{N}$  is countably infinite.
- (ii) Let  $\mathbb{Z}$  denote the set of integers. Define  $f: \mathbb{N} \rightarrow \mathbb{Z}$  by

$$f(n) = \begin{cases} \frac{n-1}{2}, & \text{if } n \text{ is odd} \\ \frac{n}{2}, & \text{if } n \text{ is even} \end{cases}.$$

Clearly  $f: \mathbb{N} \rightarrow \mathbb{Z}$  is one-one and onto. Therefore, using (i) above and Proposition 1.1 (vii),  $\mathbb{Z}$  is countably infinite. Now using Proposition 1.1 (ii) it follows that any subset of  $\mathbb{Z}$  is countable.

- (iii) Using the fact that  $\mathbb{N}$  is countably infinite and Proposition 1.1 (xiv) it is straight forward to show that  $\mathbb{Q}$  (the set of rational numbers) is countably infinite.
- (iv) Define  $f: \mathbb{R} \rightarrow \mathbb{R}$  and  $g: \mathbb{R} \rightarrow (0, 1)$  by  $f(x) = x, x \in \mathbb{R}$ , and  $g(x) = \frac{1}{1+e^x}, x \in \mathbb{R}$ . Then  $f: \mathbb{R} \rightarrow \mathbb{R}$  and  $g: \mathbb{R} \rightarrow (0, 1)$  are one-one and onto functions. It follows that  $\mathbb{R}$  and  $(0, 1)$  are continuum (using Proposition 1.1 (vii)). Further, for  $-\infty < a < b < \infty$ , let  $h(x) = (b-a)x + a, x \in (0, 1)$ . Clearly  $h: (0,1) \rightarrow (a, b)$  is one-one and onto. Again using proposition 1.1 (vii) it follows that any interval  $(a, b)$  is continuum. ■

It is clear that it may not be possible to assign probabilities in a way that applies to every situation. In the modern approach to probability theory one does not bother about how probabilities are assigned. Assignment of probabilities to various subsets of the sample space  $\Omega$  that is consistent with intuitively appealing properties (i)-(iii) of classical (or

relative frequency) method is done through *probability modeling*. In advanced courses on probability theory it is shown that in many situations (especially when the sample space  $\Omega$  is continuum) it is not possible to assign probabilities to all subsets of  $\Omega$  such that properties (i)-(iii) of classical (or relative frequency) method are satisfied. Therefore probabilities are assigned to only certain types of subsets of  $\Omega$ .