Acknowledgements

We wish to thank Sh. Sunder Singh and Gopal Krishan for word processing and secretarial assistance.

UNIT 5 BASIC CONCEPTS OF PROBABILITY THEORY

STRUCTURE

5.1 Introduction
- Objectives

5.2 Elements of Probability Theory
- Basic Terminology
- Elementary Combinatorics
- Fundamental Theorems

5.3 Random Variables
- Expectation and Variance
- Covariance and Correlation Coefficient

5.4 Summary

5.5 Terminal Questions

5.6 Solutions and Answers

5.1 INTRODUCTION

In your school curriculum you must have learnt that the concept of probability is involved in situations where the final outcome of an event cannot be predicted with absolute certainty. That is, the number of possible outcomes is more than one. For example, when we loss up a 'fair coin - as in a cricket match for determining which side will bat first - we are in doubt about the final outcome because the coin may turn up a head or a tail. Is the same true about the motion of a freely falling stone? No, in this case the final outcome - the stone will hit the ground - is not in doubt. Now suppose we roll a dice. We cannot predict which one of the six faces will appear on top. In fact, the number of possible outcomes is six and which face will turn depends on what we call chance. You can think of many other such exainplcs from everyday experience. That is why it is said, and rightly so, that if English and Mathematics are two world languages, the theory of probability is the language of mathematics since it is so intimately connected with day-to-day experience. It is, therefore, important to have a clear understanding of the theory of probability, which in a sense serves as a mathematical tool to model glorious uncertainties of chance phenomena.

To develop a mathematical theory of probability, we require some tools to communicate the outcome of an experiment. Some simple examples of an experiment might be tossing up a fair coin, drawing cards from a well shuffled pack, pre-election opinion poll and launching of a missile and observing its velocity at regular intervals of time. This leads us to the basic concepts of sample space and evenb. We hope that you are familiar with these concepts from your earlier classes. However, for completeness we have briefly discussed these in Sec. 5.2 This section also contains a short discussion of the theorems of total and compound probabilities. We conclude this section by discussing an interesting application of conditional probabilities due to Thomas Bayes.

In probability theory, the word experiment is used to describe any process that generates a set of data.

In Sec. 5.3 we have introduced the concept of random variable. Next we discuss discrete and continuous probability density functions. You will realise that the mean value (or expectation) of a random variable does not give us any idea about the spread or dispersion of a distribution. To get this information we should have a knowledge of variance, covariance and correlation coefficient. We have introduced these concepts towards the end of this unit. In the next unit we have discussed different types of continuous distributions.

On going through this unit, you may feel for taking a deep plunge into the subject. If you feel really interested, you can study the elective course on Probability and Statistics (MTE- 11).

Objectives

On studying this unit, you should be able to

- apply the fundamental theorems of probability theory to simple problems
- compute the expectation and variance of a random variable, and
- calculate covariance and correlation coefficient.

5.2 ELEMENTS OF PROBABILITY THEORY

Like any other mathematical theory, the theory of probability has some basic tools and terminology. As mentioned before, you may be familiar with them but they still find place here just for the sake of completeness. We begin by quickly recaptulating the basic terminology.

5.2.1 Basic Terminology

Suppose we toss a coin. The possible outcomes are

$$(H) \text{ and } (T)$$

where H and T correspond to "head" and "tail", respectively. You must note that we have ignored the possibility of the coin standing on its edge or its rolling away.

Let us repeat the experiment with two coins. The set of all possible outcomes may be written as

$$(H, H), (H, T), (T, H), (T, T)$$

The set of all possible outcomes of an experiment is called the **sample space** of the experiment.

Each outcome in a **sample space** is known as the sample point. The sample. space is represented by the greek symbol omega (Ω). For the first experiment, the sample space has two sample poults:

$$\Omega = \{H, T\}$$

whereas for the second experiment, the sample space has four sample points or elements:

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

It is important to emphasize here that the number of elements of a sample space depends on the information we are seeking. To illustrate this, suppose that we are interested in the number of tails which show up when two coins are tossed. Obviously, there are three possibilities: both coins may show up a tail, only one may show up a tail and neither coin turns up a tail. Then the sample space

$$\Omega = \{2, 1, 0\}$$

has three elements.

Similarly, when we toss a dice and wish to know the number that shows on the top, the sample space would be

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

But if we are interested only in knowing whether the number is even or odd, the sample space is simply

$$\Omega = \{\text{even, odd}\}$$

You may now like to work out an SAQ.

SAQ 1

Spend 2 min

A coin is flipped. If head appears, it is flipped again. Otherwise, a dice is tossed once. List the sample space.

You may now ask: Is the number of sample points always finite? If not, then how do we describe the sample space? To discover the answer to these questions, let us consider the following experiment.

A radio-active substance (^{238}U, ^{232}Th, ^{239}Pu) emits α, β and γ radiations. The number of such particles recorded by a counter may be greater than some specified number. In such cases, the sample space is written as

$$\Omega = \{ x \mid x = 0, 1, 2 \ldots \}$$

In words, it reads "Ω is the set of all x such that x takes all non-negative integers". The vertical bar is read "such that". Similarly, if we roll a pair of dice, we write

$$\Omega = \{ (i, j) \mid i = 1, 2, \ldots 6; j = 1, 2, \ldots 6 \}$$

How will you represent the set of points (x, y) on the boundary of a circle centred at origin and of radius 2? This is written as

$$\Omega = \{(x, y) \mid x^2 + y^2 = 4\}$$

We would like to mention here that a sample space containing a finite number of points is called a **discrete sample space**.

Proceeding further we recall a fact of experience that instead of the individual outcomes, we may be interested in the occurrence of particular **events** in any experiment, For instance, let us consider that a dice is rolled. We may be interested in the event E that the outcome is divisible by three. This will occur if the outcome is an element of the subset E = {3, 6} of the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$. Similarly, when two coins are tossed simultaneously, some' of the possible events are

$E_1 = \{(H, H)\}$ = Two heads appear

$E_2 = \{(H, H), (H, T), (T, H)\}$ = at least one head appears

$E_3 = \{(T, T)\}$ = No head appears

This means that to each event we can assign a collection of sample points which constitute a sub-set of the sample space. We can therefore say that

> **An event** is a subset of a sample space.

Events such as E_1 and E_3, which contain only one sample point, constitute what we call a **simple event.** Can we not say that Ω is a **sure event?** Obviously, we can.

The **null event,** denoted by ϕ, contains no element at all. A familiar situation defining a null event can be in a biological experiment where, you are asked to detect a microscopic organism by unaided eye.

Now let us consider an experiment in which the smoking habits of the employees of IGNOU are recorded. A possible sample space might be

Ω = {(non-smokers), (frequent smokers), (heavy smokers))

If the subset of smokers is some event, non-smokers correspond to a different event, which is complement of the set of smokers. We can therefore say that

> The **complement** of an event E with respect to Ω is the set of all elements of the latter that is not in former.

Obviously, null event is the complement of Ω, viz.

$$\phi = \Omega^c$$

To test your understanding we wish that you should attempt the following SAQ.

Spend 5 min

SAQ 2

List all possible events of the sample space for two simultaneously tossed coins.

You must be familiar with the basic operations of union and intersection of sets from the earlier classes. Let us now use these to define new events. You will observe that the events so formed will be subsets of the same sample space as the given events. For instance, in the tossing of a dice, we let E_1 to be the event that an even number occurs and E_2 to be the event that a number greater than three shows. Then

$$E_1 = \{2, 4, 6\}$$

and

$$E_2 = \{4, 5, 6\}$$

are subsets of the same sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$. The event $E = \{4, 6\}$ which contains elements common to E_1 and E_2 is just the intersection of E_1 and E_2. Mathematically, we write

$$E = E_1 \cap E_2$$

The intersection of two events is the event containing all elements common to them.

It readily follows that if there are no sample points common to E_1 and E_2, then

$$E_1 \cap E_2 = \phi$$

Then E_1 and E_2 are said to be disjoint or **mutually exclusive.** Stated more formally, two events are said to be mutually exclusive if they have no elements in common. You can easily see that simple events are mutually exclusive.

Quite often we are interested in the occurrence of at least one of the two events associated with an experiment. Thus, in the dice-rolling experiment, if $E_1 = \{1, 3, 5\}$ and $E_2 = \{3, 4, 5\}$, we may be interested in the occurrence of either E_1 or E_2 or both. Such an event, called the **union of** E_1 and E_2, is possible if the outcome is an element of the subset $\{1,3,4,5\}$. It is denoted by the symbol $E_1 \cup E_2$. It readily follows that Ω can be expressed as the union of all distinct simple events.

Probability of an Event

So far we have introduced the concepts of sample space and the family of events. Another very important concept is that of the probability of an event. For our purpose it is sufficient to consider only those situations for which the sample space contains a finite number of elements and all outcomes are equally likely. The likelihood of the occurrence of an event resulting from such an experiment is evaluated by means of a set of real numbers called probabilities ranging from 0 to 1. The **probability of an event A** is defined as

$$P(A) = \frac{n(A)}{n(\Omega)} \qquad (5.1)$$

where $n(\Omega)$ signifies different and equally likely outcomes out of which $n(A)$ correspond to event A.

To illustrate it, let us consider that a fair dice is rolled and calculate the probability that an even number appears. You will recall that the sample space for this experiment is

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

so that $n(\Omega) = 6$. Since the dice is fair, all these outcomes are equally likely to occur and if A represents the event that an even number appears, then

$$n(A) = 3$$

Hence,

$$P(A) = \frac{3}{6} = \frac{1}{2}$$

We would now like you to calculate the probability that (a) a multiple of three appears (b) an even number and a multiple of three appears and (c) an even number or a multiple of three appears. You will find that in the first case $P(A) = 2/6 = 1/3$, in the second case $P(B) = 1/6$, whereas in the third case $P(C) = 4/6 = 2/3$.

In solving problems in probability theory, we frequently use simple results from combinatorial analysis. We understand that you have studied these at the school level. So we shall just summarise these for ready reference.

5.2.2 Elementary Combinatorics

Let us consider two sets defined by

$$A = \{a_1, n2, \ldots, a_m\} \text{ and } B = \{b_1, b_2, \ldots, b_n\}$$

where m signifies the number of ways in which an operation A can be performed and n signifies the number of ways in which B can be performed for each of these. Suppose we pick up one element, say a_i from A and one element, say b_j from B. These constitute what is usually called a sample (a_i, b_j). Now the question arises: How many such samples are possible? The answer is obviously mn. This result is known as the multiplication rule and may be stated as follows:

If an operation can be performed in m ways and a second independent operation can be performed in n ways for each of them ways, the two operations can be performed together in mn ways.

To illustrate it, let us consider the following example.

Example 1

Calculate the number of ways in which five persons can sit in a row.

Solution

For the first person we have five possible positions in the sequence. Having chosen one of these, we find that the second person can occupy any of the remaining four positions. Proceeding in this way, you can easily calculate that the total number of ways is given by $5 \times 4 \times 3 \times 2 \times 1 = 120$.

When we have to deal with a large number of objects, say molecules of a gas in a container, the computation of the number of different ways in which some ordering/arrangements can be made may become quite cumbersome. The number of different arrangements, called **permutations**, however, can be calculated using compact mathematical relations. Let us now learn to do so.

Permutations

Consider three letters a, b, and c. The possible permutations are abc, acb, cab, cba, bca, and *bac*. You can arrive at this result using the multiplication rule, without listing different orders. There are $n_1 = 3$ choices for the first position. For each of these, there are $n_2 = 2$ choices for the second position and for each value of n_1 and n_2, $n_3 = 1$. Hence, the total number of choices is $n_1 n_2 n_3 = (3)(2)(1) = 6$. In general, n distinct objects can be arranged in n(n − 1) (n − 2) .... (3) (2) (1) ways. You may recall that this product is denoted by the symbol n !, which is read "n factorial". So we can conclude that

> A permutation is an arrangement of all or a part of a set of objects and the number of permutations of n distinct objects is n!

To generalise this result, let us consider the set

$$B = \{b_1, b_2, \ldots, b_n\}$$

Suppose we form samples of size 2, (b_i, b_j) by choosing elements from the set B. What is the number of such samples? It is easy to check that this number is n^2, provided

(i) repetition of elements is allowed, i.e., samples like (b_1, b_1), (b_2, b_2), ... are included in the count.

(ii) samples (b_i, b_j) and (b_j, b_i) for $i = j$ are counted different.

By induction, for samples of size r, the total number of such samples is n^r

This is an important result and can be stated as follows:

The number of ways of drawing a group of r objects (red balls in a box) out of a total of n objects (balls of red, blue and green colour) when repetition (replacement of balls drawn once) is permitted is n^r.

However, if samples made up of identical elements are ignored, tbe number of samples of size two will be $n(n-1)$. From this we can say that the nun . - of samples of size r without repetition is

$$n(n-1)\ldots(n-r+1) = \frac{n!}{(n-r)!} \qquad (5.2)$$

This is called the number of permutations of n different objects taken r at a time. We represent this by the symbol $^nP_r\, [= n!/(n-r)!\,]$.

So far wc have considered permutations of distinct objects. That is, all the objects were distinguishable. (This is analogous lo molecules of a classical gas.) But if we identify that the letters b and c are equal to x, then the six permutations of the letters a, b, c become axx, a n , xax, xxa, xxa, and xax. Of these, only three are distinct. Therefore, with three letters, two being the same, we have $3!/2! = 3$ distinct permutations. With four distinct letters, you have 24 different permutations. But if we let a = b = x and c = d = y, we can list only $xxyy$, $xyxy$, $yxxy$, y y n, $xyyx$ and $yxyx$ so that we have $4!/(2!)(2!) = 6$ distinct permutations. Thus

> The number of permutations of n objects of which n_1, are of one kind, n_2 are of second kind, ... n_k of kth kind is
>
> $$\frac{n!}{n_1!\, n_2!\, n_3! \,.....\, n_k!} \qquad (5.3)$$
>
> In short fonn, we express this as $\binom{n}{n_1, n_2, \ldots, n_k}$

To illustrate, let us consider that seven foreign scientists are participating in an international conference on open teaching learning organised by commonwealth of Learning at IGNOU. The hostel where arrangements for their stay are being made has one triple and two double rooms. In how many ways can we assign these rooms to them? The number is given by

$$\frac{7!}{3!\,2!\,2!} = 210$$

In many problems we are interested in the number of ways of selecting r objects from n without regard to order. These selections are called combinations, The number of such combinations is $\frac{n!}{r!\,(n-r)!}$. It is denoted by $\binom{n}{r,\, n-r}$ or simply $\binom{r}{n}$

We may, tberefore, conclude that

> The number olcombinations of n objects taken r at a time is
>
> $$^nC_r = \binom{n}{r} = \frac{n!}{r!\,(n-r)!} \qquad (5.4)$$

To fix up your ideas, you should go through tbe following example.

Example 2

IGNOU proposes to form a few committees out of its academic faculty to ensure the progress of its B.Sc. programme. To begin with, it is decided to put mathematics and physics faculty members. If two mathematicians and one physicist is to be nominated, how many committees can be formed if the faculty members of physics and mathematics are 4 and 5, respectively.

Solution

The number of ways of selecting two mathematicians out of five available is

$$\binom{5}{2} = \frac{5!}{2!\,3!} = 10$$

The number of ways of selecting one physicist out of four is

$$\binom{4}{1} = \frac{4!}{1!\,3!} = 4$$

Using the multiplication rule with $n_1 = 10$ and $n_2 = 4$, we find that the number of committees that can be formed is

$$n = n_1 n_2 = 10 \times 4 = 40$$

You will learn that it is often easier to calculate the probability of some event from known probabilities of other events. This may also be true when the event in question can be represented as the union of two other events or as the complement of other events. Some important laws and theorems, which frequently simplify the computation of probabilities in complex situations, can be used. Let us now learn about these.

5.2.3 Fundamental Theorems

The first theorem deals with union of events and is called the **theorem of total probability** or **the additive law.** For simplicity, let us first consider two mutually exclusive events E and F. The probability for one of these to occur is the sum of the probabilities of separate events. Mathematically, if $E \cap F = \phi$, we write

$$P(E \cup F) = P(E) + P(F) \qquad (5.5)$$

You can prove it using Eq. (5.1) for the simple case of finite sample spaces. Since $E \cap F = \phi$, we have

$$n(E \cup F) = n(E) + n(F)$$

so that

$$P(E \cup F) = \frac{n(E \cup F)}{n(\Omega)} = \frac{n(E) + n(F)}{n(\Omega)}$$

$$= \frac{n(E)}{n(\Omega)} + \frac{n(F)}{n(\Omega)} = P(E) + P(F) \qquad (5.6)$$

You can readily extend this result to n mutually exclusive events $E_1, E_2, \ldots E_n$:

$$P(E_1 \cup E_2 \cup E_3 \ldots \cup E_n) = P(E_1) + P(E_2) + \ldots + P(E_n) \qquad (5.7a)$$

In summation form, we can write

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) \qquad (5.7b)$$

The result contained in Eq. (5.5) can be generalised by relaxing the constraint $E \cap F = \phi$. That is, we can consider any two events. This is usually stated as the theorem of total probabilities:

> **Theorem of Total Probability:** If E and F are any two events, then (5.8)
> $$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

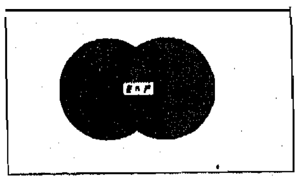

Fig. 5.1: Venn diagram representing E and E ∪ F

To prove this theorem, consider the diagram shown in Fig. 5.1. (This is known as Venn diagram.) You will note that the quantity appearing on LHS of Eq. (5.8), $P(E \cup F)$ is the sum of the probabilities of sample points in $E \cup F$. Now $P(E) + P(F)$ is the sum of all probabilities in E plus the sum of all probabilities in F. This incorporates the probabilities in $E \cap F$ twice. Therefore, we must subtract this probability once to obtain $P(E \cup F)$.

To illustrate this theorem we have given a solved example. You should go through it carefully.

In Venn diagram, we represent the sample space by a rectangle and the events by circles drawn inside it. This diagram pictorially depicts the relation between events and the corresponding sample space.

Example 3

A number is chosen at random from the first twenty five natural numbers. Calculate the probability that the integer is divisible by 2 or 3.

Solution

Let E be the event that the number selected is divisible by 2 and F signifies that the number is divisible by 3. So we have $n(E) = 12$, $n(F) = 8$. Of these, four numbers are common. So $n(E \cap F) = 4$. Hence,

$$P(E) = \frac{12}{15}, P(F) = \frac{8}{25} \text{ and } P(E \cap F) = \frac{4}{25}$$

$$\therefore\ P(E \cup F) = \frac{16}{25}$$

You can now check your progress by attempting the following SAQ. We hope that you will gain more insight into probability and its interpretation.

Spend 15 min

SAQ 3

i) Prove Eq. (5.8) mathematically.

ii) Compute $P(E \cup F \cup G)$.

iii) Calculate the probability of getting a total of 7 or 11 when a pair of dice are tossed:

Let us again consider rolling of a dice. Suppose you are told that the number which shows up is even or odd. Then this information will influence the computation of probability that a six will appear. Let us now proceed to learn to compute the probability of an event occurring even though it is known that some other event, which can affect it, has occurred. This is called the **conditional probability** and is denoted by $P(F|E)$. In words, this symbol is read as " the probability of F given that E occurs."

The conditional probability of F, given E, if $P(E) > 0$ is defined by

$$P(F \mid E) = \frac{P(E \cap F)}{P(E)} \tag{5.9}$$

Before we prove this result, let us consider an illustration.

Example 4

Two firms manufacture similar voltmeters. Finn A produces 1,000 voltmeters out of which 30 are defective. Firm B produces 4,500 out of which 100 are defective. A product is chosen at random and found to be defective. What is the probability that it belongs to firm B?

Solution

Let the event E denote that the ammeter chosen is defective and the event F signify that it came from B. Then we compute $P(F|E)$ using the relation

$$P(F \mid E) = \frac{P(E \cap F)}{P(E)}$$

$$\frac{(100/5500)}{(130/5500)}$$

$$= \frac{10}{13}$$

To check whether the added knowledge of voltmeter being defective affects the probability that it came from B or not, let us compute $P(F) = \frac{n(F)}{N(\Omega)} = \frac{4500}{5500} = \frac{9}{11}$. This is greater than $P(F \mid E)$.

As another illustration, consider the following example.

Example 5

The probability that a regularly scheduled flight departs on time is $P(D) = 0.83$, the probability that it arrives on time is $P(A) = 0.82$, and the probability that it departs and arrives on time is $P(D \cap A) = 0.78$. Calculate the probability that a plane (i) arrives, on time if it departed on time and (ii) departs on time if it arrived on time.

Solution

(i) The probability that the plane arrives on time, given that it departed on time, is

$$P(A \mid D) = \frac{P(D \cap A)}{P(D)}$$

$$= \frac{0.78}{0.83}$$

$$= 0.94$$

(ii) The probability that the plane departs on time, given that it had arrived on time., is

$$P(D \mid A) = \frac{P(D \cap A)}{P(A)}$$

$$= \frac{0.78}{0.82}$$

$$= 0.95$$

You may now like to check your progress by solving the following SAQ.

SAQ 4

Spend 10 *min*

A small town has 500 male and 400 female adults of which 460 males and 140 females are employed. One of these individuals is to be selected at random to advertise setting up of new industries in the town. Calculate the probability of choosing a man, given that he is employed.

We now know that conditional probability allows for an alteration of the probability of an event in the light of given information. But in some situations, it is possible that

$$P(E \mid F) = P(E) \text{ for } P(F) \neq 0 \qquad (5.10)$$

That is, the occurrence of F has no influence on the occurrence of E. Then, E is said to be independent of F. For this special case you can readily show that

$$P(E \cap F) = P(E)P(F) \qquad (5.11)$$

This is illustrated the following example.

Example 6

Assume that 3 coins are tossed. Let E be the event that a tail appears on the first coin, and F be the event that a head appears on the third throw. Are E and Findependent?

Solution

We have here Ω = (HHH), (HTH), (HHT), (HTT), (THH), (TTH), (THT), (TTT)

Obviously $n(\Omega) = 8\ (= 2^3)$, $n(E) = 4$, $N(F) = 4$, and $n(E \cap F) = 2$, Hence,

$$P(E) = 1/2, P(F) = 1/2, \text{ and } P(E \cap F) = 1/4$$

Since Eq. (5.11) is satisfied, E and F are independent.

An important application of conditional probabilities was made by the English philosopher, Thomas Bayes. To illustrate this let us reconsider SAQ 4. Suppose that you are now told that 36 of those employed and 12 of those unemployed are members of a social organisation: Scientists' Forum for Peace. We now wish to know: Is the individual selected, a member of this organisation? To this end, let, us suppose that the probability of this event is A. Then, by referring to Fig. 5.2, we can write A as the union of two mutually exclusive events $E \cap A$ and $E^c \cap A$:

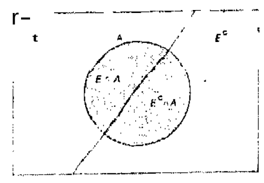

Fig. 5.2 Venn diagramm for the events A, E, and E^c

$$A = (E \cap A) \cup (E^c \cap A)$$

Using Eq. (5.5) we can write

$$P(A) = P[(E \cap A) + (E^c \cap A)]$$

$$= P(E \cap A) + P(E^c \cap A)$$

On combining this result with Eq. (5.7) we find that

$$P(A) = P(E)\,P(A \mid E) + P(E^c)\,P(A \mid E^c)$$

In SAQ 4, the total number of employed persons is 600 and the number of unemployed persons is 300. Therefore,

$$P(E) = \frac{600}{900} = \frac{2}{3},\ P(A \mid E) = \frac{36}{900} = \frac{3}{50}$$

$$P(E^c) = \frac{300}{900} = \frac{1}{3} \text{ and } P(A \mid E^c) = \frac{12}{300} = \frac{1}{25}.$$

Hence,

$$P(A) = \left[\frac{2}{3}\right]\left[\frac{3}{50}\right] + \left[\frac{1}{3}\right]\left[\frac{1}{25}\right]$$

$$= \frac{4}{75}$$

A generalization of this illustration to the case where the sample space is partitioned into k sub-sets is stated as the theorem of **total** probability. Before we do so, it is important to recall the definition of the partition of a set Ω. The family of sets $\{E_1, E_2, \ldots. E_n\}$ is called a partition of Ω if

(a) $E_i \subset \Omega$ for $i = 1, 2, \ldots, n$

(b) $E_i \cap E_j = \phi$ for $i, j = 1, 2, \ldots n;\ i \neq j$

and

(c) $E_1 \cup E_2 \cup \ldots \cup E_n = \bigcup_{i=1}^{n} E_i = \Omega$ (5.12)

> **Generalised** Theorem of **Total** Probability
>
> If $\{E_1, E_2, \ldots E_n\}$ constitute a partition of the **sample space** Ω such that $P(E_i) \neq 0$ for $i = 1, 2, \ldots, n$, then for **any event** A of Ω
>
> $$P(A) = \sum_{i=1}^{n} P(E_i)\,P(A \mid E_i) = \sum_{i=1}^{n} P(E_i \cap A) \quad (5.13)$$

To prove this theorem we note that by **definition**

$$\Omega = E_1 \cup E_2 \cup \ldots \cup E_n$$

Hence,

$$A = A \cap \Omega = A \cap (E_1 \cup E_2 \cup \ldots \cup E_n)$$

$$= (A \cap E_1) \cup (A \cap E_2) \cup \ldots \cup (A \cap E_n)$$

It is also clear that

$$(A \cap E_i) \cap (A \cap E_j) = \phi \text{ for } i \neq j$$

because $A \cap E_i \subset E_i$ and by definition E_i's are disjoint. Therefore,

$$P(A) = \sum_{i=1}^{n} P(A \cap E_i) = \sum_{i=1}^{n} P(E_i)\, P(A \mid E_i) \qquad (5.14)$$

It is now a straightforward exercise to establish Bayes' theorem.

Bayes' Theorem: If the events $E_1, E_2, \ldots E_n$ constitute a partition of the sample space P, where $P(E_i) \neq 0$ for i = 1, 2, ... n, then for any event A in Ω such that $P(A) \neq 0$

$$P(E_i \mid A) = \frac{P(E_i \cap A)}{\sum_{i=1}^{n} P(E_i \cap A)} = \frac{P(E_i)\, P(A \mid E_i)}{\sum_{i=1}^{n} P(E_i)\, P(A \mid E_i)} \qquad (5.15)$$

fori = 1, 2, ..., n.

Proof: By the definition of conditional probability

$$P(E_i \mid A) = \frac{P\,(E_i \cap A)}{P(A)}$$

Now we use the theorem of total probability in the denominator to obtain the required result:

$$P(E_i \mid A) = \frac{P(E_i \cap A)}{\sum_{i=i}^{n} P(E_i \cap A)}$$

We can put Bayes' theorem in a somewhat more useful form using the multiplicative rule $[P(E_i \cap A) = P(E_i)\, Y(A \mid Ei)]$ for the occurrence of any two events in both numerator and denominator:

$$P(E_i \mid A) = \frac{P(E_i)\, P(A \mid Ei)}{\sum_{i=1}^{n} P\,(E_i)\, P(A \mid E_i)}$$

This theorem is also called the formula for probabilities of hypotheses because the events $E_1, E_2, \ldots, E_n$ may be considered as hypotheses to account for the occurrence of the event A. These ideas are illustrated in the following example.

Example 7

A transformer producing company has four assembly plants which are working with efficiency of 30%, 20%, 40% and 10%, respectively. The probnbilities that a transformer produced by these plants is defective are 0.02, 0.03, 0.04 and 0.01, respectively. If a transformer is chosen at random, what is the probability that it is defective? If a transformer, chosen at random is found to be defective, what is the probability that it came from the fourth plant?

Solution

Let E be the event that transformer chosen is defective and the event E_i signify that it came from plant i (i = 1, 2, 3, 4). From the theorem of total probability, we have

$$P(E) = \sum_{i=1} P(E_i)\, P(E \mid E_i)$$

$$= (0.3)\,(0.02) + (0.2)\,(0.03) + (0.4)\,(0.03) + (0.1)\,(0.04)$$

$$= 0.028$$

From Eq, (5.15) it follows that

$$P(E_4 \mid E) = \frac{P(E_4)\, P(E \mid E_4)}{0.028} = \frac{0.004}{0.028} = \frac{1}{7}$$

So far we have introduced the basic theorems of conditional and total probabilities. One of the basic concepts of probability theory is that of a **random variable.** We shall now introduce this concept.

5.3 RANDOM VARIABLES

Two children are playing a game by tossing a fair coin. The sample space for the experiment is $\Omega = \{H, T\}$. The child calling head accrues one unit but loses one unit if a tail appears. You can express this correspondence as

$$H \Leftrightarrow 1 \quad \text{and} \quad T \Leftrightarrow -1$$

We can mathematically express this result by introducing a real-valued random variable X which takes the value 1 and –1, with probability 1/2 in each case. That is,

$$P(X = 1) = 1/2, \; P(X = -1) = 1/2$$

Let us now roll a dice. If X denotes the number appearing on top, its possible values are 1, 2, 3, 4, 5, 6 with probability 116 in each case. Similarly, let us suppose that two coins are tossed. If X denotes the number of heads which appear, then X can assume numerical values of 2,1 or 0 with probabilities 1/4, 1/2 and 1/4, respectively. These numbers are random quantities determined by the outcome of the experiment. These should bc viewed as values assumed by the random variable X. So we can say that

> A random variable is a function that associates a real number with each element in sample space Ω.

If ω is a point in Ω and X a random variable, then $X(\omega)$ is the value of the random variable at ω. Usually, we are concerned neither with X nor 8. Instead, we wish to know the probability that the value of the random variable is in a certain stated. This is denoted by $P(X \in A)$.

A random variable is said to be a discrete random variable if it takes only a finite number of values, i.e., the set of possible outcomes is countable. When a random variable takes on countless values, it is called a continuous random variable. For example, the position of a point in the interval [0,1] is a continuous variable. In most practical problems, continuous random variables represent measured data such as temperature, lifetime, height, etc., On the other hand, discrete random variables represent count data such as number of casualties in an accident, number of defective items in a given sample, etc.

We now know that a discrete random variable assumes each of its values with a certain probability. Frequently, it is convenient to represent all the probabilities of a random variable by a formula. Such a formula would necessarily be a function of the numerical values, which we denote by $f(x_i)$ such that

$$f(x_i) = P(x = x_i), \qquad i = 1, 2, 3, \ldots \qquad (5.16a)$$

The set of ordered pairs $(x_i, f(x_i))$ is called the probability function or probability distribution of the discrete random variable X, if for each possible outcome xi,

(i) $f(x_i) \geq 0$ (5.16b)

(ii) $\sum_i f(x_i) = 1$ (5.16c)

For the continuous random variable, we must have

$$P(a \leq x \leq b) = \int_a^b f(x)\, dx \qquad (5.17a)$$

$$f(x) \geq 0 \text{ for all } x \qquad (5.17b)$$

and

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \qquad (5.17c)$$

There are several interesting examples of continuous distributions. In Unit 7 of this block you will learn that statistical errors in an experiment follow the normal distribution. The speeds of molecules in a gas have Maxwellian distribution. (Fig. 5.3a). The lifetimes of radioactive nuclei or the free paths in a gas exhibit exponential distribution (Fig. 5.3b).

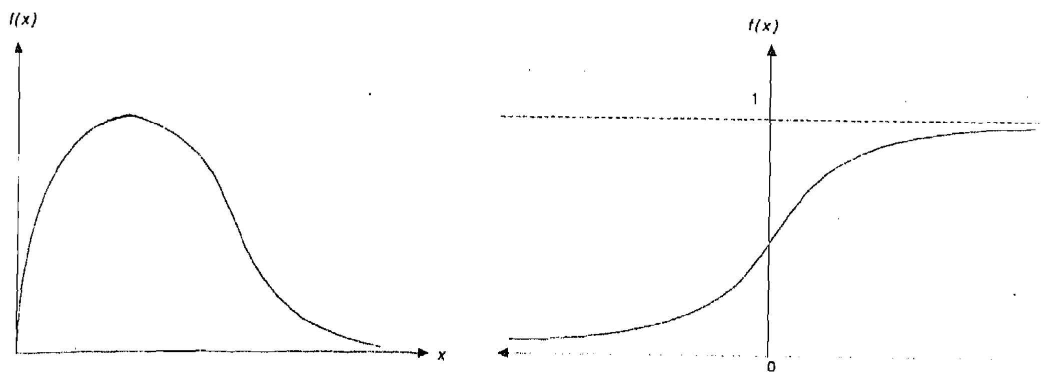

Fig. 5.3: A few typical continuous distributions

We shall now define the expectation of a random variable. It is of special importance in statistics because it gives us an idea as to where the probability distribution is centred. It, however, does not give us quantitative estimate of the spread of distribution (or dispersion of values) about the mean. We should, therefore, characterise the variability in the distribution. The most important measure of variability of a random variable is referred to as the **variance.** Let us learn about these in detail now.

5.3.1 Expectation and Variance

Let us assume that a random variable X assumes values $x_1, x_2, x_3, ...$ with corresponding probabilities $f(x_1), f(x_2), f(x_3), ...$ The mathematical expectation or mean of X is defined by

$$E(X) = \langle X \rangle = \sum x_i f(x_i) \tag{5.18}$$

provided $\Sigma x_i f(x_i) < \infty$, i.e., the series converges absolutely and X is discrete.

If X is continuous, the corresponding definition is

$$E(X) = \langle X \rangle = \int_{-\infty}^{\infty} x f(x)\, dx \tag{5.19}$$

To aypreciatc the idea involved in this concept, you should go through the following worked out examples.

Example 8

Suppose that X denotes the number obtained on rolling a fair dice. Calculate $E(X)$.

Solution

We know that X can take values from 1 to 6 with probability 1/6 in each case. Hence,

$$E(X) = (1)(1/6) + (2)(1/6) + (3)(1/6) + (4)(1/6) + (5)(1/6) + (6)(1/6)$$

$$3.5$$

Example 9

For the hydrogen atom, $f(\mathbf{r})$ is given by

$$f(\mathbf{r})\,\mathbf{dr} = \frac{1}{\pi a_0^3} \exp\left(-\frac{2r}{a_0}\right) dr$$

where a_0 is the first Bohr radius and r is the distance between the electron and proton. Calculate $E(r)$.

Solution

You will note that $f(r)\,dr$ signifies continuous distribution. So using Eq. (5.19) we can write

$$E(r) = \frac{1}{\pi a_0^3}\int_0^\infty r \exp\left(-\frac{2r}{a_0}\right) dr$$

From Unit 3 of Block 1 of this course you will recall that $dr = 4\pi r^2 dr$. Inserting this result in the above expression, we can write

$$E(r) = \frac{4}{a_0^3}\int_0^\infty r^3 \exp\left(-\frac{2r}{a_0}\right) dr$$

The gamma function is defined by

$$\Gamma(n+1) = \int_0^\infty e^{-x} x^n\, dx = n\ \Gamma(n)$$

An important property of gamma function is that

$$\Gamma(n) = (n-1)\,\Gamma(n-1)$$

with $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(1) = 1$.

To evaluate this integral we introduce a change of variable by defining

$$\frac{2r}{a_0} = x$$

so that $dr = (a_0/2)\,dx$ and $r^3 dr = (a_0/2)^4 x^3 dx$. On inserting these results, we get

$$E(x) = \frac{a_0}{4}\int_0^\infty \exp(-x)\, x^3 dx$$

$$= \frac{a_0}{4}\,\Gamma(4)$$

$$= 6\, a_0$$

where $\Gamma(n)$ is gamma function.

In the language of probability theory, the expectation of x is known as the first moment of the distribution about the origin.

As pointed out earlier, to obtain more precise information about a distribution, we define the variance of X. It is denoted by σ_X^2 or $\mathrm{Var}(X)$. When X is discrete, the variance of X is given by

$$\sigma_X^2 = \mathrm{Var}(X) = \sum_x [x - E(X)]^2 f(x) \qquad (5.20a)$$

When X is continuous we have

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - E(X))^2\, f(x)\, dx \qquad (5.20b)$$

The positive square root of the variance is called the standard deviation of X. This property enables us to rewrite Eq. (5.20a) in an alternative and often preferred form. To do so, we recall that for the discrete case we can write

$$\mathrm{Var}(X) = \sum_x [x - E(X)]^2 f(x) \qquad (5.21)$$

$$= \sum_x x^2 f(x) - 2E(X)\sum_x x f(x) + [E(X)]^2 \sum_x f(x)$$

$$= E(X^2) - [E(X)]^2 \qquad (5.22)$$

since $\sum_x x f(x) = E(X)$ and $\sum_x f(x) = 1$.

We would now like you to prove this result when X is continuous. You have to repeat the above procedure step by step, with summations replaced by integrations.

Let us reconsider Example 8. We know that

$E(X) = 3.5$

The values of $x_i - E(X)$ are given by the set

$$\{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$$

Hence, Eq. (5.20a) gives

$$\begin{aligned}\text{Var}(X) &= 2[(2.5)^2 \times (1/6) + (1.5)^2 \times (1/6) + (0.5)^2 \times (1/6)] \\ &= (1/3) \times [6.25 + 2.25 + 0.251 \\ &= (1/3) \times (8.75) = 2.92\end{aligned}$$

The variance of X is known as the **second moment** of the distribution about the mean (x) and gives us detailed information about the probability distribution.

We would now like you to do the following SAQ.

SAQ 5

Spend 10 min

The weekly demand for a popular cold drink is a continuous random variable X having the probability

$$f(x) \begin{cases} 2(x-1) & 1 < x < 2 \\ & \text{elsewhere} \end{cases}$$

Calculate the mean and variance.

5.3.2 Covariance and Correlation Coefficient

The variance or standard deviation has meaning only when we compare two or more distributions having the same units of measurement. And we can comment upon the uniformity or variability of the distribution. That is, it is not meaningful to compare the variance of a distribution of heights to the variance of a distribution to aptitude scores. That is why we have so far confined ourselves to univariate distributions. However, we frequently come across situations in which more than one variable is involved. Suppose we have a class of Nstudents and record the height, x, and weight, y, of each student. This is an example of a **bivariate distribution.** To each student there corresponds a pair of values (x, y) of the variates. Suppose the pair occurs jointly f times. Then, obviously

$$\sum_i f_i = N \qquad (5.23)$$

and the corresponding probabilities $f(x_i, y_i)$ are given by f_i/N.

The **covariance** between random discrete variables X and Y, denoted by Cov(X, Y) is defined as

$$\text{Cov}(X, Y) = \sum_i f(x_i, y_i)\ (x_i - \langle X \rangle)(y_i - \langle Y \rangle) \qquad (5.24)$$

This important notion was introduced by Galton in 1885 and is very useful in the study of nature of association between two variatcs. You will get a flavour of this in Unit 7. An alternative, and more often used, formula for covariance of two random variables X and Y with means $\bar{X}$ and $\bar{Y}$, respectively is given by

$$\text{Cov}(X, Y) = E(XY) - \bar{X}\bar{Y} \qquad (5.25)$$

To prove this result we note that

$$\begin{aligned}\text{Cov}(X, Y) &= \sum_x \sum_y (x - \bar{X})(y - \bar{Y}) f(x, y) \\ &= \sum_x \sum_y (xy - \bar{X}y - x\bar{Y} + \bar{X}\bar{Y}) f(x, y)\end{aligned}$$

$$= \sum_x \sum_y xy\, f(x,y) - \overline{X} \sum_x \sum_y y\, f(x,y)$$

$$- \overline{Y} \sum_x \sum_y x\, f(x,y) + \overline{X}\,\overline{Y} \sum_x \sum_y f(x,y)$$

If we recall the definitions of $\overline{X}$, and $\overline{Y}$ and note that $\sum \sum f(x, y) = 1$, the desired result follows readily.

The counterpart of Eq. *(5.24)* when X and Y are continuous is given by

$$\mathrm{Cov}(X, Y) = \int\int f(x, y)\,(x - \langle X \rangle)\,(y - \langle Y \rangle)\, dx\, dy \qquad (5.26)$$

You can prove this result proceeding along lines used to arrive at Eq. (5.25).

If X and Y are independent, it is straightforward to show that

$$E(XY) = E(X)\,E(Y) \qquad (5.27)$$

Then, it readily follows from Eqs. *(5.25)* and *(5.27)* that

$$\mathrm{Cov}\,(X, Y) = 0 \qquad (5.28)$$

Although the covariance between two random variables gives information about the nature of their relationship, its magnitude does not suggest anything about the strength of the relationship, since Cov (X, *Y*) is not scale free. However, we can determine the strength of such relationship by means of **correlation** coefficient, r_{XY}. It is defined as

$$r_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}\,(X, Y)}{\sigma_X\,\sigma_Y} \qquad (5.29)$$

This important notion was introduced by Karl Pearson in 1895. If *X* and Y are independent, Eq. (5.28) implies that

$$r_{XY} = 0 \qquad (5.30)$$

However, you must be careful not to conclude the converse. It may very well happen that $r_{XY} = 0$ and still X and *Y* may not be independent.

If the pair (x_i, y_i) occurs f_i times, it is easy to show that

$$r_{XY} = \frac{\sum_i f_i x_i y_i - \frac{1}{N}\left(\sum_i f_i x_i\right)\left(\sum_i f_i y_i\right)}{\sqrt{\sum_i f_i x_i^2 - \frac{1}{N}\left(\sum_i f_i x_i\right)^2} \times \sqrt{\sum_i f_i y_i^2 - \frac{1}{N}\left(\sum_i f_i y_i^2\right)^2}}$$

where $\sum_i f_i = N$.

This relation is particularly useful in practical calculations of correlation coefficient. You will learn to use it in detail in Unit *7*.

We are now ready to summarise the important points of this unit.

5.4 SUMMARY

- The probability of an event *A* is defined as

$$P(A) = \frac{n(A)}{n(\Omega)}$$

- $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
  If $E \cap F = \phi$, this reduces to $P(E \cup F) = P(E) + P(F)$. This is known as the **theorem** of total probability.

- $P(E \cap F) = P(E)\,P(E \mid E) = P(F)\,P(E \mid F)$
  This is called the **theorem** of compound probability.

- **Bayes'** Theorem Let $\{E_1, ..., E_n\}$ be a partition of Ω, where $P(E_i) \neq 0$, $i = 1, 2, ..., n$. Let E be any event with $P(E) \neq 0$. Then, we have

$$P(E_j \mid E) = \frac{P(E_j)\, P(E \mid E_j)}{\sum_{i=1}^{n} P(E_i)\, P(E \mid E_i)}$$

- A random variable is said lo be discrete if it assumes only a finite number of values. A continuous random variable assumes a non-denumerable number of values.

- The expectation of X is defined as

$$E(X) = (X) = \begin{cases} \sum_i x_i f(x_i) & X \text{ discrete} \\ \int_{-\infty}^{\infty} x f(x)\, dx & X \text{ continuous} \end{cases}$$

- The variance of X is given by

$$\text{Var}(X) = \sigma_X^2 = E(X^2) - [E(X)]^2$$

- The covariance of two random variables X and Y is defined as

$$\text{Cov}(X, Y) = E(XY) - \overline{X}\overline{Y}$$

- The correlation coefficient is given by

$$r_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\, \text{Var}(Y)}}$$

If X and Y are independent, $r_{XY} = 0$

5.5 TERMINAL QUESTIONS

Spend 40 min

1. The probability that A will pass an examination is 3/7 and the probability that B will pass it is 4/7. What is the probability that at least one of them passes the examination?
2. An ordinary dice is tossed. If A is the event of 'getting an odd number' and B is the event of 'getting a perfect square' calculate $P(B \mid A)$.
3. In a well-shuffled deck of cards, what is the probability that the Queen of spades is next to the Jack of spades?
4. Suppose four coins are tossed. Let X designate the number of heads which appear. Calculate $E(X)$.

5.6 SOLUTIONS AND ANSWERS

SAQs

1. $\{HH, HT, T1, T2, T3, T4, T5, T6\}$
2. All possible events for the sample space

$\{(H, H), (H, T), (T, H), (T, T)\}$ are :

ϕ, $\{(H, H)\}$, $\{(H, T)\}$ $\{(T, H)\}$ $\{(T, T)\}$

$\{(H, H), (H, T)\}$, $\{(H, H), (T, H)\}$, $\{(H, H), (T, T)\}$

$\{(H, T), (T, H)\}$, $\{(H, T), (T, T)\}$, $\{(T, H), (T, T)\}$

$\{(H, H), (H, T), (T, H)\}$, $\{(H, H), (H, T), (T, T)\}$

$\{(H, T), (T, H), (T, T)\}, \{(H, H), (T, H), (T, T)\}, \Omega$

In general, for a sample space of size n, the total number of events is 2^n.

3. (i) From Fig. 5.1, it is clear that if E and F are mutually exclusive events, then

$$E \cup F = E \cup (E^c \cap F) \quad \text{.. (i)}$$

and $$F = (E \cap F) \cup (E^c \cap F) \quad \text{.... (ii)}$$

Using Eq. (5.5), we have

$$P(E \cup F) = P(E) + P(E^c \cap F)$$

and $$P(F) = P(E \cap F) + P(E^c \cap F)$$

On combining these results we find that

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

which is the required result.

(ii) We wish to compute $P(E \cup F \cup G)$.

Let $F \cup G = H$; then

$$P(E \cup F \cup G) = P(E \cup H) = P(E) + P(H) - P(E \cap H)$$

Also $$P(H) = P(F) + P(G) - P(F \cap G)$$

$$P(E \cap H) = P(E \cap F \cup G)$$
$$= P[(E \cap F) \cup (E \cap G)]$$

Hence,

$$P(E \cap H) = P(E \cap F) + P(E \cap G) - P[(E \cap F) \cap (E \cap G)]$$
$$= P(E \cap F) + P(E \cap G) - P(E \cap F \cap G)$$

Therefore, we have

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) - P(F \cap G)$$
$$- P(G \cap E) + P(E \cap F \cap G)$$

(iii) Let A be the event that 7 occurs and B the event that 11 comes up. Now a total of 7 occurs for 6 of the 36 sample points but a total of 11 occurs for only 2 of the sample points. Since all sample points are equally likely, we have $P(A) = 1/6$ and $P(B) = 1/18$. Since 7 or 11 can be summed up on the same toss, we can say that the events A and B are mutually exclusive. Therefore,

$$P(A \cup B) = P(A) + P(B)$$
$$= (1/6) + (1/18)$$
$$= 2/9$$

4. Suppose that M signifies the event that a man is chosen and E signifies that the chosen one is employed. Since the total number of emyloyed persons is 600, we find that

$$P(M \mid E) = \frac{460}{600} = \frac{23}{30}$$

5. $$E(X) = \int_1^2 x f(x)\, dx = 2 \int_1^2 x(x-1)\, dx = \frac{5}{3}$$

and

$$E(X^2) = 2 \int_1^2 x^2(x-1)\, dx = \frac{17}{6}$$

From Eq, (5.22) it follows that

$$\text{Var}(X) = \frac{17}{6} - \frac{25}{9} = \frac{1}{18}$$

TQs

1. The required probability is

$$P(A \cup B) = P(A) + P(B) - P(A)P(B)$$

since A and B are independent. Hence,

$$P(A \cup B) = \frac{3}{7} + \frac{4}{7} - \frac{12}{49} = \frac{37}{49}$$

2. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$. We have

$$A = \{1, 3, 5\}, B = \{1, 4\} A \cap B = \{1\}$$

Hence,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{(1/6)}{(3/6)} = \frac{1}{3}$$

3. Let $E \equiv$ Queen and Jack of spades arc neighbouring cards

Ω = {ordered 52-tuples denoting all possible arrangements of the 52 cards}

Introduce the hypothesis,

$E_1 \equiv$ the Queen of spades occupies the top position in the deck

$E_2 \equiv$ the Queen of spades occupies the bottom position in the deck

$E_3 \equiv$ the Queen of spades is somewhere within the deck.

Obviously, $P(E_1) = P(E_2) = 51!/52! = 1/52$, $P(E_3) = 50/52 = 25/26$. Since only one card is next to the top or bottom cards and two cards are next to a card within the deck, we have

$$P(E \mid E_1) = P(E \mid E_2) = 1/51, \; P(E \mid E_3) = 2/51$$

Then

$$P(E) = P(E_1)P(E \mid E_1) + P(E_2)P(E \mid E_2) + P(E_3)P(E \mid E_3)$$
$$= 1/26$$

4. We have

$$P(X = 0) = 1/16, P(X = 1) = 1/4, P(X = 2) = 3/8$$
$$P(X = 3) = 1/4, P(X = 4) = 1/16$$
$$E(X) = 0 \times \left(\frac{1}{16}\right) + 1 \times \left(\frac{1}{4}\right) + 2 \times \left(\frac{3}{8}\right) + 3 \times \left(\frac{1}{4}\right) + 4 \times \left(\frac{1}{16}\right)$$
$$= \frac{1}{4} + \frac{3}{4} + \frac{3}{4} + \frac{1}{4} = 2$$