UNIT 7

EXPECTATION AND MOMENTS GENERATING FUNCTION

Structure

7.1 INTRODUCTION

In Units 4 and 5, you studied discrete and continuous univariate random variables and their different functions like CDF, PMF, PDF, etc. In Unit 6, you have studied bivariate discrete and continuous random variables and their corresponding functions. Most of the time, we are interested in summary statistics of a probability distribution like mean, variance, etc. You know that the mean of a distribution is a very popular measure of central tendency. In probability theory mean is known as the expected value of a probability distribution. In fact, the expected value is a weighted mean where weights are probabilities. In Sec. 7.2, you will study to obtain the expected value of a probability distribution. Expected value alone does not tell all the information of our interest so we need more measures of the distribution. In Sec. 7.3, you will study a function known as moment generating function. With the help of moment generating function, we can obtain many measures of the distribution some of them are discussed in Sec. 7.3.

What we have discussed in this unit is summarised in Sec. 7.4. Self-Assessment Questions (SAQs) have been given in some sections which are generally based on the content discussed in that section. But to give you a good practice of what we have discussed in this unit some more questions based on the entire unit are given in Sec. 7.5 under the heading Terminal Questions. Due to the reason mentioned in Sec. 1.1 of Unit 1 of this course, solutions of all the SAQs and Terminal Questions are given in Sec. 7.6.

In the next unit, you will study transformations of univariate and bivariate random variables.

Expected Learning Outcomes

After completing this unit, you should be able to:

- obtain the expected value of a probability distribution; and
- define moment generating function and find a particular moment of your interest for a given probability distribution.

7.2 EXPECTED VALUE OF A PROBABILITY DISTRIBUTION AND ITS PROPERTIES

You know how to calculate the simple arithmetic mean (AM) of given data and what is interpretation of the AM. For example, suppose you are interested in getting AM of the outcomes of a fair die. The outcomes of a fair die are 1, 2, 3, 4, 5 and 6. So, from school mathematics, you know that

$$AM = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = 3.5 \qquad \ldots (7.1)$$

To interpret AM geometrically, assume that each outcome is a box of equal shape and size and placed at points 1, 2, 3, 4, 5, and 6 on real line refer to Fig. 7.1. If you try to balance the line containing these six boxes then it will balance at point 3.5 shown by an arrow. So, AM is interpreted as the balancing point in the present case on the real line between 1 and 6. Fig. 7.1 suggests that the balancing point is 3.5 which is shown in the figure by an arrow.

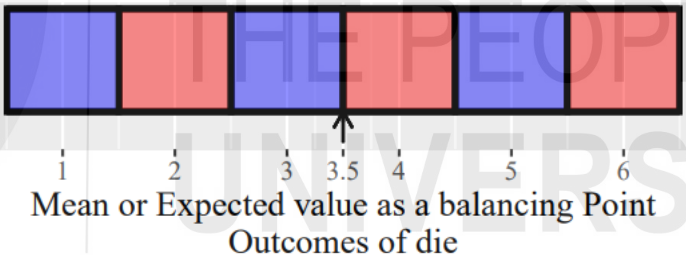

Fig. 7.1: Visualisation of AM 3.5 of 1, 2, 3, 4, 5 and 6 as a balancing point

Equation (7.1) can also be written as

$$AM = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1) + \frac{1}{n}(x_2) + \frac{1}{n}(x_3) + \ldots + \frac{1}{n}(x_n) \qquad \ldots (7.2)$$

$$= w_1 x_1 + w_2 x_2 + \ldots + w_n x_n = \sum_{i=1}^{n} w_i x_i \qquad \ldots (7.3)$$

where each $w_i = \frac{1}{n}$ is known as weight of i^{th} observation

In the example of the outcomes of a fair die, we have

$$AM = \frac{1}{6}(1) + \frac{1}{6}(2) + \frac{1}{6}(3) + \frac{1}{6}(4) + \frac{1}{6}(5) + \frac{1}{6}(6) = \frac{21}{6} = 3.5 \qquad \ldots (7.4)$$

Here, each $w_i = \frac{1}{6},\ i = 1, 2, \ldots, 6.$ When we use weights then AM is known as weighted mean (WM). To understand this important point let us suppose that your marks in the assignment and Term End Exam (TEE) of this course are 60 and 90 respectively. Also, suppose that the marks of your friend in the assignment and TEE of this course are 90 and 60 respectively. As per the present evaluation methodology, the weightage of assignment and TEE are 30% and 70% respectively. So, as per the evaluation methodology, your and your friend final marks in MST-012 will be calculated as follows:

$$\text{Your marks } = 0.3(60) + 0.7(90) \qquad \left[\begin{array}{l} \because \text{ assignment weight} = 30\% = 0.3 \\ \text{and TEE weight} = 70\% = 0.7 \end{array}\right]$$

$$= 18 + 63 = 81$$

$$\text{Your friend marks } = 0.3(90) + 0.7(60) \quad [\text{Same reason}]$$

$$= 27 + 42 = 69$$

If you calculate AM of marks then AM of both of you will be 75.

$$\left[\begin{array}{l} \because \text{ AM is a particular case of WM where each weight} = \dfrac{1}{\text{No of observations}} \\ \text{Here we have two observations so each } w_i = \dfrac{1}{2} = 0.5. \\ \text{So AM} = 0.5(60) + 0.5(90) = 30 + 45 = 75 \end{array}\right]$$

From the above discussion, you noticed the effect of weights on AM. Also, observe that the sum of all weights is 1. In the case of die $\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{6}{6}=1$ and in the case of your marks in MST-012 it is 0.3 + 0.7 = 1. … (7.5)

So, if you have data with weights of the observations then you can easily obtain the weighted mean of the given data using (7.3). In probability theory, WM is known as the **expected value**. But in probability theory, you either have

- probability mass function (PMF) if random variable X is discrete or
- probability density function (PDF) if random variable X is continuous.

So, a natural question that may arise in your mind is, how the expected value is obtained when we have PMF or PDF instead of weighted data. In this section, you will learn to obtain expected value using PMF or PDF. First, we will learn to calculate the expected value in probability theory and then discuss its properties.

7.2.1 Expected Value of a Probability Distribution

Probability distribution may be discrete or continuous. So, we have to further classify our discussion under these two subheadings.

Expected Value of a Discrete Probability Distribution

Equation (7.5) gives a hint to tackle the problem at hand. The common things between weights and probabilities are:

- both are non-negative, i.e., weights ≥ 0 as well as probabilities ≥ 0, and
- their sum is 1, i.e., the sum of all weights = 1 as well as the sum of all probabilities = 1. … (7.6)

So, if we replace weights with corresponding probabilities we are done. So, the expected value of a discrete distribution can be defined as follows.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and X be a discrete random variable defined on the sample space Ω which can attain either a finite number of values $x_1, x_2, x_3, \ldots, x_n$ or a countable infinite number of values $x_1, x_2, x_3, \ldots$ then the **expected value or expectation** of X is denoted by E(X) and is defined by

$$E(X) = \sum_{x} \underset{\substack{\uparrow \\ \text{Possible value} \\ \text{of rv X}}}{x} \; \underset{\substack{\uparrow \\ \text{Probability of} \\ \text{possible value}}}{\mathcal{P}(X = x)}, \quad \text{provided it converges absolutely} \qquad \ldots (7.7)$$

We have discussed absolutely and conditionally convergent of an infinite series in Unit 4 of the course MST-011. Here absolutely convergent means

$$E(X) = \sum_{x} \left| x\,\mathcal{P}(X = x) \right| < \infty. \qquad \ldots (7.8)$$

If (7.8) holds then we say that random variable X or corresponding probability distribution has finite expectation. If (7.7) produces $+\infty$ or $-\infty$, then we say that the random variable X or corresponding probability distribution has an infinite expectation. But if the series in (7.7) oscillates finitely or infinitely then we say that the expectation of the random variable X does not exist. What is the meaning of oscillates finitely or infinitely, we have discussed this in Unit 4 of the course MST-011. … (7.9)

In particular, if we want to write (7.7) specifically in the case X attains a finite number of values $x_1, x_2, x_3, \ldots, x_n$ then it can be written as

$$E(X) = \sum_{i=1}^{n} \underset{\substack{\uparrow \\ \text{Possible value} \\ \text{of rv X}}}{x_i} \; \underset{\substack{\uparrow \\ \text{Probability of} \\ \text{possible value}}}{\mathcal{P}(X = x_i)} \qquad \ldots (7.10)$$

and if we want to write (7.7) specifically in the case X attains a countably infinite number of values $x_1, x_2, x_3, \ldots$ then it can be written as

$$E(X) = \sum_{n=1}^{\infty} \underset{\substack{\uparrow \\ \text{Possible value} \\ \text{of rv X}}}{x_n} \; \underset{\substack{\uparrow \\ \text{Probability of} \\ \text{possible value}}}{\mathcal{P}(X = x_n)} \qquad \ldots (7.11)$$

Another way of writing (7.7) is

$$E(X) = \sum_{\omega \in \Omega} \underset{\substack{\uparrow \\ \text{Value of rv X corresponding} \\ \text{to sample point } \omega}}{X(\omega)} \; \underset{\substack{\uparrow \\ \text{Probability of the} \\ \text{singleton event } \{\omega\}}}{\mathcal{P}(\{\omega\})} \qquad \ldots (7.12)$$

Equation (7.12) is a very basic result and will help us in proving some results.

In general, expected value of a function g(X) of the random variable X is given by

$$E(g(X)) = \sum_{\omega \in \Omega} \underset{\substack{\uparrow \\ \text{Value of g(X) corresponding} \\ \text{to sample point } \omega}}{g(X(\omega))} \; \underset{\substack{\uparrow \\ \text{Probability of the} \\ \text{singleton event } \{\omega\}}}{\mathcal{P}(\{\omega\})} \qquad \ldots (7.13)$$

Now we consider one example so that you can understand how simple it is to apply the formula of expected value. Let us consider the random experiment of throwing a tetrahedral die discussed in Example 3 in Unit 4 of this course. Recall that the sample space for this example is

$$\Omega = \{1, 2, 3, 4\}^2 = \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (2, 4), (3, 1), (3, 2), (3, 3), (3, 4), (4, 1), (4, 2), (4, 3), (4, 4)\}.$$

Also recall that in this example, X denotes the maximum of the two outcomes. So, X can take values 1, 2, 3 and 4 and all the 16 possible outcomes of this random experiment are equally likely, so appropriate probability measure on Ω is uniform probability measure. So,

$$\mathcal{P}\left(\{\omega_i\}\right) = \frac{1}{16} \;\; \forall \; \omega_i \in \Omega, \; i = 1, 2, 3, \ldots, 16.$$

In Unit 4, we have already obtained probability mass function of X which is given by Table 7.1 as follows.

Table 7.1: Probability mass function or probability distribution of rv X

X	1	2	3	4
$p_X(x)$	1/16	3/16	5/16	7/16

Now, using (7.10) expected value of probability distribution of X or simply of X is given by

$$E(X) = \sum_{i=1}^{4} x_i \, \mathcal{P}(X = x_i) = (1)\left(\frac{1}{16}\right) + (2)\left(\frac{3}{16}\right) + (3)\left(\frac{5}{16}\right) + (4)\left(\frac{7}{16}\right)$$

$$= \frac{1 + 6 + 15 + 28}{16} = \frac{50}{16} = \frac{25}{8} = 3.125 \qquad \ldots (7.14)$$

If we are interested in expected value of $g(X) = 3X^2 + 5X + 2$ then it is given by

$$E(g(X)) = \sum_{i=1}^{4} g(x_i)\,\mathcal{P}(X = x_i) = \left[3(1)^2 + 5(1) + 2\right]\left(\frac{1}{16}\right) + \left[3(2)^2 + 5(2) + 2\right]\left(\frac{3}{16}\right)$$

$$+ \left[3(3)^2 + 5(3) + 2\right]\left(\frac{5}{16}\right) + \left[3(4)^2 + 5(4) + 2\right]\left(\frac{7}{16}\right)$$

$$= \frac{10 + 72 + 220 + 490}{16} = \frac{792}{16} = \frac{99}{2} = 49.5 \ldots (7.15)$$

Now we discuss how to calculate expected value for continuous distribution and what modification is required compared to discrete distributions.

Expected Value of a Continuous Probability Distribution

In Unit 5 of the course MST-011 and in Units 5 and 6 of this course, we have discussed that in continuous world job of summation is done by integration and in what way. So, if f_X be the PDF of continuous random variable X then expected value of X is given by

$$E(X) = \int_{-\infty}^{\infty} \overset{\text{Value of } X}{\overset{\downarrow}{x}} \; \underbrace{\overset{\text{Density at } x}{\overset{\downarrow}{f_X(x)}} \; \overset{\text{length of small interval around } x}{\overset{\downarrow}{dx}}}_{\text{Product of density } f_X(x) \text{ at } x \text{ and length } dx \text{ around } x \text{ is the probability that X lies in } \left(x - \frac{dx}{2},\, x + \frac{dx}{2}\right) \text{ refer (5.28a) of this course}} \qquad \ldots (7.16)$$

provided this integral converges absolutely. An integral $\int_a^b f(x)\,dx$ is said to be absolutely convergent if the integral $\int_a^b |f(x)|\,dx$ converges, i.e., if $\int_a^b |f(x)|\,dx < \infty$.

In general, expected value of a function g(X) of the random variable X is given by

$$E(g(X)) = \int_{-\infty}^{\infty} \underset{\text{at } X=x}{\overset{\text{Value of } g(X)}{\downarrow}} g(x) \; \underbrace{\overset{\text{Density at x}}{\downarrow} f_X(x) \quad \overset{\text{length of small interval around x}}{\downarrow} dx}_{\text{Product of density } f_X(x) \text{ at } x \text{ and length } dx \text{ around } x \text{ is the probability that X lies in } \left(x-\frac{dx}{2},\, x+\frac{dx}{2}\right) \text{ refer (5.28a) of this course}} \qquad \ldots (7.17)$$

For example, consider the continuous probability distribution discussed in Example 2 of Unit 5 of this course. Using (7.16) expected value of X for the continuous probability distribution of X is given by

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_0^1 x\big(6x(1-x)\big)dx = 6\int_0^1 \left(x^2 - x^3\right)dx = 6\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1$$

$$= 6\left[\frac{1}{3} - \frac{1}{4} - 0\right] = 6\left(\frac{4-3}{12}\right) = \frac{1}{2} = 0.5 \qquad \ldots (7.18)$$

If we are interested in expected value of $g(X) = 3X^2 + 5X + 2$ then it is given by

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx = \int_0^1 \left(3x^2 + 5x + 2\right)\big(6x(1-x)\big)dx$$

$$= 6\int_0^1 \left(3x^3 + 5x^2 + 2x - 3x^4 - 5x^3 - 2x^2\right)dx = 6\int_0^1 \left(-3x^4 - 2x^3 + 3x^2 + 2x\right)dx$$

$$= 6\left[\frac{-3x^5}{5} - \frac{2x^4}{4} + x^3 + x^2\right]_0^1 = 6\left[\frac{-3}{5} - \frac{1}{2} + 1 + 1 - 0\right]_0^1 = 6\left(\frac{-6-5+20}{10}\right) = \frac{27}{5}$$

$\ldots (7.19)$

After understanding how to calculate expected value for discrete as well as continuous cases, we now discuss some properties of expectation in the next subsection of this section.

7.2.2 Properties of Expectation

You have studied how to find expected value of a discrete and continuous random variable in the previous subsection of this section. In this subsection we will discuss some properties of expectation which hold for both discrete and continuous random variables. So, let us state and prove these properties as follows.

Statement: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and X, Y be two random variables defined on the sample space Ω then expectation satisfies the following properties.

(a) E(aX) = a E(X) … (7.20p1)

(b) E(X + Y) = E(X) + E(Y) … (7.20p2)

In general, $E\left(X_1 + X_2 + X_3 + \ldots + X_n\right) = E\left(X_1\right) + E\left(X_2\right) + E\left(X_3\right) + \ldots + E\left(X_n\right)$

… (7.20p3)

(c) E(a) = a, where a is a constant … (7.20p4)

(d) If $X(\omega) \geq 0 \;\; \forall \;\; \omega \in \Omega,$ then E(X) ≥ 0 … (7.20p5)

(e) If $X(\omega) \geq Y(\omega) \;\; \forall \;\; \omega \in \Omega,$ then E(X) ≥ E(Y) … (7.20p6)

(f) $\left|E(X)\right| \leq E(\left|X\right|)$ for any random variable X, provided expectations exist … (7.20p7)

Proof: We shall prove these results for discrete and continuous cases separately.

Discrete case:

(a) By definition of expectation refer (7.12), we have

$$\begin{aligned} E(aX) &= \sum_{\omega\in\Omega} (aX)(\omega)\mathcal{P}(\{\omega\}) = \sum_{\omega\in\Omega} aX(\omega)\mathcal{P}(\{\omega\}) && \left[\begin{matrix}\text{Using (2.24) of Unit 2 of}\\ \text{the course MST-011}\end{matrix}\right] \\ &= a\sum_{\omega\in\Omega} X(\omega)\mathcal{P}(\{\omega\}) && [\because \text{a is independent of } \omega] \\ &= aE(X) && [\text{Using (7.12)}] \end{aligned}$$

(b) By definition of expectation refer (7.12), we have

$$E(X) = \sum_{\omega\in\Omega} X(\omega)\mathcal{P}(\{\omega\}) \quad \text{… (7.21)}$$

and $$E(Y) = \sum_{\omega\in\Omega} Y(\omega)\mathcal{P}(\{\omega\}) \quad \text{… (7.22)}$$

Adding (7.21) and (7.22), we get

$$\begin{aligned} E(X)+E(Y) &= \sum_{\omega\in\Omega} X(\omega)\mathcal{P}(\{\omega\}) + \sum_{\omega\in\Omega} Y(\omega)\mathcal{P}(\{\omega\}) \\ &= \sum_{\omega\in\Omega} (X(\omega)+Y(\omega))\mathcal{P}(\{\omega\}) \\ &= \sum_{\omega\in\Omega} (X+Y)(\omega)\mathcal{P}(\{\omega\}) && \left[\begin{matrix}\text{Using (2.19) of Unit 2 of}\\ \text{the course MST-011}\end{matrix}\right] \\ &= E(X+Y) && [\text{Using (7.12) for random variable } X+Y] \end{aligned}$$

(c) You know expected value is a property of probability distribution so, whenever you are interested in getting expected value of a constant, ‘a’ (say), it means you are considering a distribution of a variable which always attains single value ‘a’. In the world of probability theory, it means, we are dealing with a random variable J (say) which attains value ‘a’ with probability 1. That is $\mathcal{P}(J=a)=1.$ But total probability is 1 which has been assigned to the event J = a so PMF of the random variable J can be written as

$$p_J(j) = \begin{cases} 1, & \text{if } j=a \\ 0, & \text{otherwise} \end{cases}$$

Now, by definition of expectation refer (7.12), we have

$$\begin{aligned} E(a) &= E(J) && [\because \text{J attains value a with probability 1. So, } J=a \text{ always}] \\ &= a\mathcal{P}(J=a) \\ &= a(1) = a && [\because \mathcal{P}(J=a)=1] \end{aligned}$$

(d) By definition of expectation refer (7.12), we have

$$E(X)=\sum_{\omega\in\Omega}X(\omega)\mathcal{P}\left(\{\omega\}\right) \quad \dots (7.23)$$

We are given that $X(\omega)\geq 0 \ \forall \ \omega\in\Omega$ … (7.24)

Also, we know that probability of an event is always ≥ 0. So,
$\mathcal{P}\left(\{\omega\}\right)\geq 0 \ \forall \ \omega\in\Omega$ … (7.25)

Combining (7.24) and (7.25), we have

$$X(\omega)\mathcal{P}\left(\{\omega\}\right)\geq 0 \ \forall \ \omega\in\Omega$$

$$\Rightarrow \sum_{\omega\in\Omega}X(\omega)\mathcal{P}\left(\{\omega\}\right)\geq 0$$

$$\Rightarrow E(X)\geq 0 \qquad [\text{Using (7.23)}] \quad \dots (7.26)$$

(e) We are given that $X(\omega)\geq Y(\omega) \ \forall \ \omega\in\Omega$.

$$\Rightarrow X(\omega)-Y(\omega)\geq 0 \ \forall \ \omega\in\Omega.$$

$$\Rightarrow \left(X-Y\right)(\omega)\geq 0 \ \forall \ \omega\in\Omega.$$

$$\Rightarrow E\left(X-Y\right)\geq 0 \qquad [\text{Using (7.26)}]$$

$$\Rightarrow E\left(X+(-Y)\right)\geq 0$$

$$\Rightarrow E(X)+E(-Y)\geq 0 \qquad [\text{Using (7.20p2)}]$$

$$\Rightarrow E(X)-E(Y)\geq 0 \qquad [\text{Using (7.20p1), where } a=-1]$$

$$\Rightarrow E(X)\geq E(Y)$$

(f) We know that for any random variable X, we always have $X\leq|X|$

$$\Rightarrow E\left(X\right)\leq E\left(|X|\right) \qquad [\text{Using (7.20p6)}] \quad \dots (7.27)$$

Also, $-X\leq|X|$, so again using (7.20p6), we have

$$E\left(-X\right)\leq E\left(|X|\right)$$

$$\Rightarrow -E\left(X\right)\leq E\left(|X|\right) \quad \dots (7.28)$$

Combining (7.27) and (7.28), we have

$$\left|E\left(X\right)\right|\leq E\left(|X|\right)$$

Continuous case: Like discrete case, we can prove all properties for continuous case also. We have to just replace summation by integration and PMF by PDF times dx. Let us prove first property.

(a) If f_X be the probability density function of the random variable X then using (7.16), we have

$$E(aX)=\int_{-\infty}^{\infty}(ax)f_X(x)\,dx=a\int_{-\infty}^{\infty}x\,f_X(x)\,dx$$

$$=aE(X) \qquad [\text{Using (7.16)}]$$

Similarly, other properties can be proved in continuous case.

Now we discuss two important expectation theorems known as addition and multiplication theorems.

Addition Theorem of Expectation: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and X, Y be two random variables defined on the sample space Ω then addition theorem of expectation says that E(X + Y) = E(X) + E(Y). … (7.29)

In general, if we have n random variables $X_1, X_2, X_3, \ldots, X_n$ defined on the sample space Ω then

$$E\left(X_1 + X_2 + X_3 + \ldots + X_n\right) = E\left(X_1\right) + E\left(X_2\right) + E\left(X_3\right) + \ldots + E\left(X_n\right) \qquad \ldots (7.30)$$

Further, if we have a countable sequence of random variables $X_1, X_2, X_3, \ldots$ defined on the sample space Ω then

$$E\left(X_1 + X_2 + X_3 + \ldots\right) = E\left(X_1\right) + E\left(X_2\right) + E\left(X_3\right) + \ldots \qquad \ldots (7.31)$$

Proof: We have already proved (7.29) refer (7.20p2). (7.30) can be proved by repeating (7.29) as many times as required. (7.31) can be proved using principle of mathematical induction.

Multiplication Theorem of Expectation: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and X, Y be two independent random variables defined on the sample space Ω then multiplication theorem of expectation says that

E(XY) = E(X)E(Y). … (7.32)

Proof: If S_X and S_Y denote supports of random variables X and Y then independence of random variables X and Y implies

$$\left.\begin{aligned} &\mathcal{P}\left(X = x,\ Y = y\right) = \mathcal{P}\left(X = x\right)\mathcal{P}\left(Y = y\right) && \forall\ x \in S_X,\ y \in S_Y \\ &f_{X,\,Y}\left(x,\ y\right) = f_X\left(x\right)f_Y\left(y\right) && \forall\ x \in S_X,\ y \in S_Y \end{aligned}\right\} \qquad \ldots (7.33)$$

Also, by the definition of expectation, we have

$$E(X) = \sum_{x \in S_X} x\,\mathcal{P}\left(X = x\right) \quad \text{or} \quad E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \qquad \ldots (7.34)$$

$$E(Y) = \sum_{y \in S_Y} y\,\mathcal{P}\left(Y = y\right) \quad \text{or} \quad E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy \qquad \ldots (7.35)$$

For discrete case

$$\begin{aligned} E(XY) &= \sum_{x \in S_X} \sum_{y \in S_Y} x\,y\,\mathcal{P}\left(X = x,\ Y = y\right) \\ &= \sum_{x \in S_X} \sum_{y \in S_Y} x\,y\,\mathcal{P}\left(X = x\right)\mathcal{P}\left(Y = y\right) && \left[\text{Using } (7.33)\right] \\ &= \sum_{x \in S_X} x\,\mathcal{P}\left(X = x\right) \sum_{y \in S_Y} y\,\mathcal{P}\left(Y = y\right) \\ &= E(X)E(Y) && \left[\text{Using } (7.34) \text{ and } (7.35)\right] \end{aligned}$$

For continuous case

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\,y\,f_{X,\,Y}(x,\ y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\,y\,f_X(x)\,f_Y(y)\,dx\,dy \qquad [\text{Using } (7.33)]$$

$$= \left(\int_{-\infty}^{\infty} x\ f_X(x)\,dx\right)\left(\int_{-\infty}^{\infty} y\,f_Y(y)\,dy\right)$$

$$= E(X)E(Y) \qquad [\text{Using } (7.34) \text{ and } (7.35)]$$

Let us do some examples.

Example 1: A coin is tossed once. If X denotes the number of heads and p be the probability of getting a head then find the expected value of X.

Solution: Here the random variable X can attain values 0 and 1 with probabilities 1 – p and p respectively. So, by the definition of expectation, we have

$$E(X) = \sum_{x=0}^{1} x\,\mathcal{P}\left(X = x\right) = 0\mathcal{P}\left(X = 0\right) + 1\mathcal{P}\left(X = 1\right) = 0(1-p) + 1(p) = p \quad \dots\ (7.36)$$

Example 2: Two coins are tossed once. Probability of getting a head in each coin is p. Find the expected value of number of heads in this experiment.

Solution: Let X denote the number of heads on the first coin and Y denote the number of heads on the second coin, then from Example 1, we know that

$$E(X) = p \text{ and } E(Y) = p \qquad \dots\ (7.37)$$

Since we are interested in finding the expected value of the total number of heads on both the coins, therefore, expected number of heads in this experiment is E(X + Y).

Using (7.20p2), we have

$$\begin{aligned} E(X+Y) &= E(X) + E(Y) \\ &= p + p \qquad [\text{Using } (7.37)] \\ &= 2p \qquad \dots\ (7.38) \end{aligned}$$

Example 3: A random experiment of tossing n coins simultaneously is performed. Each coin has probability p of getting a head. Find the expected value of number of heads in this experiment.

Solution: Let X_i, $i = 1, 2, 3, \dots, n$ denote the number of heads on the i^{th} coin, then from Example 1, we know that

$$E(X_1) = p,\ E(X_2) = p,\ E(X_3) = p,\ \dots,\ E(X_n) = p \qquad \dots\ (7.39)$$

So, expected number of heads in this experiment is $E(X_1 + X_2 + X_3 + \dots + X_n)$.

Using (7.20p3), we have

$$\begin{aligned} E(X_1 + X_2 + X_3 + \dots + X_n) &= E(X_1) + E(X_2) + E(X_3) + \dots + E(X_n) \\ &= p + p + p + \dots + p \qquad [\text{Using } (7.39)] \\ &= np \qquad \dots\ (7.40) \end{aligned}$$

Example 4: **The St. Petersburg Paradox Game**: You are offered to play a coin toss game. The game is like this: The game will end as soon as you get the first tail. You go on to toss the coin until you get tail. If you get tail in the:

- first flip you will get Rs 2, (first round)
- second flip you will get Rs 4, (second round)
- third flip you will get Rs 8, (third round) and so on.

In general, if game ends at n^{th} round, then you will get Rs 2^n. Suppose you accept offer of this game then find the expected value of this game. Assume that coin used for tossing is fair.

Solution: Let X denote your gain then

$$X = 2^n, \ n = 1, 2, 3, 4, \ldots \qquad \ldots (7.41)$$

where n represents number of rounds game may run

Since expectation requires the probability for each value of the random variable X so first we have to find the same for each round as follows.

$$\mathcal{P}(\text{getting tail in the first round}) = \frac{1}{2} \qquad [\because \text{coin is fair}]$$

$$\mathcal{P}(\text{getting tail in the second round})$$
$$= \mathcal{P}(\text{getting head in the first round}) \times \mathcal{P}(\text{getting tail in the second round})$$
$$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$\mathcal{P}(\text{getting tail in the third round})$$
$$= \mathcal{P}(\text{getting head in the first round}) \times \mathcal{P}(\text{getting head in the second round})$$
$$\times \mathcal{P}(\text{getting tail in the third round})$$
$$= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

and so on $\mathcal{P}(\text{getting tail in the } n^{th} \text{ round}) = \dfrac{1}{2^n}$ ... (7.42)

Using (7.41), (7.42) and definition of expected value, we have

$$E(X) = \sum_x x\mathcal{P}(X = x) = (2)\frac{1}{2} + (4)\frac{1}{4} + (8)\frac{1}{8} + \ldots = 1 + 1 + 1 + \ldots = \infty \qquad \ldots (7.43)$$

Remark 1: Equation (7.43) may surprise you because in real world no one have infinite amount of money. But crucial point to be noted here is if you toss a fair coin then as number of rounds increase probability that coin will continue producing head is decreasing half of the previous round. However, as per rule of the game amount is increasing 2 times of the amount in the previous round. This increasing and decreasing feature of two quantities mathematically dictates sum in (7.43) and force to tends to infinity. But in reality, no one can be so lucky that a fair coin continuously goes on showing head and shows first tail after a very large number of trials. This is a very interesting problem. You can have different explanation to argue on the value given by (7.43).

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

SAQ 1

Probability mass function of the random variable X is given by

$$\mathcal{P}\left(X = n\right) = \frac{1}{2^{n+1}}, \ n = 0, 1, 2, 3, \ldots$$

Find expected value of X.

7.3 MOMENT GENERATING FUNCTION

In the previous section, you have learnt to calculate expected value of a random variable or probability distribution. In this section, you will see that expected value is the first order moment of the probability distribution. This beautiful idea of expected value or expectation of X is not only gives us mean (first raw moment) of the probability distribution but also useful in obtaining variance (second central moment), covariance and other moments of the distribution. Expected value gives information about the centre of the probability distribution. But in statistics other than centre of the probability distribution, we are also interested in getting information about spread, symmetry, fatness of tails (left and right) of the distribution. To get such measures, we have to define higher order moments of the probability distribution. In this section, we will not only define higher order moments of the probability distribution but also discuss a function known as moments generating function (MGF) which generates all order moments of the probability distribution if it exists. At the end of this section, we will also discuss some important properties of MGF.

Let us first define different types of moments in general.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and X be a random variable defined on the sample space Ω then rth order (r = 1, 2, 3, 4, …)

- **raw moment** is denoted by μ_r' and is defined by $\mu_r' = E\left(X^r\right)$ … (7.44)
- **central moment** is denoted by μ_r and is defined by

$$\left.\begin{aligned} \mu_r &= E\left(X - \mu\right)^r \\ \text{or } \mu_r &= E\left(X - E(X)\right)^r \end{aligned}\right\} \quad \ldots (7.45)$$

- **standardised moment** if denoted by μ_r^{sd} and is defined by

$$\left.\begin{aligned} \mu_r^{sd} &= E\left(\frac{X - \mu}{\sigma}\right)^r \\ \text{or } \mu_r^{sd} &= E\left(\frac{X - E(X)}{\sigma}\right)^r \end{aligned}\right\} \quad \ldots (7.46)$$

where $\sigma = \text{standard deviation} = \sqrt{E\left(X - E(X)\right)^2}$ … (7.47)

Interpretation of some particular cases are listed in Table 7.2 given as follows.

Table 7.2: Interpretation of different moments up to order 4

Value of r	Raw moment	Central moment	Standardised moment
1	Mean or Average distance from 0 (if all values are +ve or all are -ve). If both signs present then distance interpretation will not work.	0 Sum of deviation about expected value is always zero. Hence, their expected value will be zero	0 Sum of deviation about expected value in standard units is always zero.
2	Average square distance from 0	Variance or Average square distance from mean	1 Variance of observations in standard units is always 1.
3	Average cubic distance from 0 (if all values are +ve or all are -ve). If both signs present then distance interpretation will not work.	0 If distribution is symmetric about mean	Skewness of the distribution $E\left(\frac{X-E(X)}{\sigma}\right)^3$
4	Average 4^{th} power distance from 0	Average 4^{th} power distance from mean	Kurtosis $E\left(\frac{X-E(X)}{\sigma}\right)^4$

Here r is taking countably infinite values so to find raw moments for each value of r one by one is not feasible and is very time-consuming task even if we find first few moments one by one. But fortunately, we have no need to do so for each value of r because we have a function known as moment generating function which takes care of this issue and is discussed as follows.

Moment Generation Function (MGF): Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and X be a random variable defined on the sample space Ω. Let $A = \left\{t \in \mathbb{R} : E\left(e^{tX}\right) \text{ exists}\right\}$ then the function $M_X : A \to \mathbb{R}$ defined by

$$M_X(t) = E\left(e^{tX}\right), \; t \in A \qquad \dots (7.48)$$

is called moment generating function for the random variable X, provided A is non empty and there exists some positive real number $\varepsilon > 0$ such that $(-\varepsilon, \varepsilon) \subseteq A$. That is, there exists an open interval $(-\varepsilon, \varepsilon)$ around 0 where $M_X(t)$ is finite. … (7.49)

Note that MGF is defined in terms of expected value and expected value is a property of a distribution. So, we will use it in Units 9 to 16 where we will discuss different discrete and continuous probability distributions. … (7.50)

For understanding purpose let us obtain MGF of a Bernoulli(p) distribution. You will study Bernoulli distribution in Unit 9 you may refer to (9.46). PMF of Bernoulli distribution which takes two values 0 and 1 is given by

$$p_X(x) = \begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0 \end{cases}$$

Or $$\mathcal{P}\left(X = x\right) = \begin{cases} p^x (1-p)^{1-x}, & \text{if } x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

By definition of MGF, we have

$$M_X(t) = E\left(e^{tX}\right) = e^{t(0)}\,\mathcal{P}\left(X = 0\right) + e^{t(1)}\,\mathcal{P}\left(X = 1\right) = 1(1-p) + e^t p = q + p\,e^t \quad \dots (7.51)$$

where q = 1 – p

Note that (7.51) is defined for all real values of t. Hence, MGF of Bernoulli(p) is defined for all real values of t.

Now, let us discuss one result which will explain why it is called moment generating function.

By definition of MGF, we have

$$M_X(t) = E\left(e^{tX}\right)$$
$$= E\left(1 + tX + \frac{t^2X^2}{\underline{|2}} + \frac{t^3X^3}{\underline{|3}} + \frac{t^4X^4}{\underline{|4}} + \dots + \frac{t^rX^r}{\underline{|r}} + \frac{t^{r+1}X^{r+1}}{\underline{|r+1}} + \dots\right)\begin{bmatrix}\text{Using (1.31)} \\ \text{of MST-011}\end{bmatrix}$$

Using (7.31) in RHS, we have

$$M_X(t) = E(1) + E(tX) + E\left(\frac{t^2X^2}{\underline{|2}}\right) + E\left(\frac{t^3X^3}{\underline{|3}}\right) + \dots + E\left(\frac{t^rX^r}{\underline{|r}}\right) + E\left(\frac{t^{r+1}X^{r+1}}{\underline{|r+1}}\right) + \dots$$

Using (7.20p1) in RHS, we have

$$M_X(t) = 1 + t\,E(X) + \frac{t^2}{\underline{|2}}E\left(X^2\right) + \frac{t^3}{\underline{|3}}E\left(X^3\right) + \dots + \frac{t^r}{\underline{|r}}E\left(X^r\right) + \frac{t^{r+1}}{\underline{|r+1}}E\left(X^{r+1}\right) + \dots$$

Using (7.44) in each term of RHS except the first term, we have

$$M_X(t) = 1 + t\,\mu_1' + \frac{t^2}{\underline{|2}}\mu_2' + \frac{t^3}{\underline{|3}}\mu_3' + \dots + \frac{t^r}{\underline{|r}}\mu_r' + \frac{t^{r+1}}{\underline{|r+1}}\mu_{r+1}' + \dots \quad \dots (7.52)$$

Note that in equation (7.52) coefficient of $\frac{t^r}{\underline{|r}}$ is μ_r'. But we know that n^th^ derivative of t^n with respect to t is $\underline{|n}$. So, if we differentiate (7.52) r times with respect to t and after that put t = 0, then we will only left with μ_r' all other terms will vanish. Terms before t^n will vanish because of higher derivatives than powers of t and terms after t^n will vanish due to putting t = 0 after differentiation. So, after differentiating successively and putting t = 0, we are getting the same thing in which we are interesting. To implement this strategy for different values of r, let us differentiate (7.52) with respect to t successively.

$$M_X^{(1)}(t) = \mu_1' + \frac{2t}{\underline{|2}}\mu_2' + \frac{3t^2}{\underline{|3}}\mu_3' + \dots + \frac{r\,t^{r-1}}{\underline{|r}}\mu_r' + \frac{(r+1)t^r}{\underline{|r+1}}\mu_{r+1}' + \dots \quad \dots (7.53)$$

$$M_X^{(2)}(t) = \frac{2}{\underline{|2}}\mu_2' + \frac{6t}{\underline{|3}}\mu_3' + \dots + \frac{r(r-1)\,t^{r-2}}{\underline{|r}}\mu_r' + \frac{(r+1)r\,t^{r-1}}{\underline{|r+1}}\mu_{r+1}' + \dots \quad \dots (7.54)$$

$$M_X^{(3)}(t) = \frac{6}{\underline{|3}}\mu_3' + \ldots + \frac{r(r-1)(r-2)t^{r-3}}{\underline{|r}}\mu_r' + \frac{(r+1)r(r-1)t^{r-2}}{\underline{|r+1}}\mu_{r+1}' + \ldots \quad \ldots (7.55)$$

$\vdots$

$$M_X^{(r)}(t) = \frac{\underline{|r}}{\underline{|r}}\mu_r' + \frac{(r+1)r(r-1)\ldots 3.2\,t}{\underline{|r+1}}\mu_{r+1}' + \ldots \quad \ldots (7.56)$$

$\vdots$

where (r) in $M_X^{(r)}(t)$ denotes r^{th} derivative of $M_X(t)$ with respect to t.

Now, putting t = 0 in (7.53) to (7.56), we get

$$M_X^{(1)}(0) = \mu_1' \quad \ldots (7.57)$$

$$M_X^{(2)}(0) = \mu_2' \quad \ldots (7.58)$$

$$M_X^{(3)}(0) = \mu_3' \quad \ldots (7.59)$$

$\vdots$

$$M_X^{(r)}(0) = \mu_r' \quad \ldots (7.60)$$

$\vdots$

So, all the raw moments in which we are interested can be obtained simply by differentiating (7.52) an appropriate number of times and then putting t = 0. This is the reason it is given the name MGF.

Another particular moment in which we will be interested throughout the course is the second central moment known as **variance** of the distribution. But in real life problems, expected value of the distribution comes out in decimals so if we use (7.45) taking r = 2 as a formula to calculate variance then calculations become difficult, time consuming and the most importantly error in final solution will be more due to lots of rounded off of values. So, let us obtain a simplified expression for variance instead of (7.45) for r = 2. To do so let us replace r by 2 in (7.45) and simplify that expression as follows.
… (7.61)

$$\text{Variance of the distribution} = \mu_2 = E\left(X - E\left(X\right)\right)^2 \quad \ldots (7.62)$$

$$= E\left(X^2 - 2XE(X) + \left(E(X)\right)^2\right) \quad \left[\begin{array}{l}\text{Using the identity}\\ (a-b)^2 = a^2 - 2ab + b^2\end{array}\right]$$

$$= E\left(X^2\right) - 2E\left(X\right)E\left(X\right) + \left(E(X)\right)^2 \quad \left[\begin{array}{l}\text{Using (7.20p1), (7.20p2) and (7.20p3)}\\ \text{as } E(X) \text{ and } \left(E(X)\right)^2 \text{ are constant}\end{array}\right]$$

$$= E\left(X^2\right) - 2\left(E(X)\right)^2 + \left(E(X)\right)^2$$

$$= E\left(X^2\right) - \left(E(X)\right)^2$$

$$\therefore \text{ Variance of the distribution} = \mu_2 = E\left(X^2\right) - \left(E(X)\right)^2 \quad \ldots (7.63)$$

In future, most of the time, we will use (7.63) to obtain variance of the distribution instead of (7.45) for r = 2.

We are denoting expectation of X by E(X). Similarly, variance of X is denoted by V(X). So, (7.62) and (7.63) respectively can be written as

$$\text{Variance of the distribution} = V(X) = \mu_2 = E\left(X - E(X)\right)^2 \quad \dots (7.64)$$

$$\therefore \text{Variance of the distribution} = \mu_2 = V(X) = E\left(X^2\right) - \left(E(X)\right)^2 \quad \dots (7.65)$$

One more common notation that we will use sometime is Cov(X, Y) for **covariance of X and Y** and it is defined by

$$\text{Cov}(X, Y) = E\left(\left(X - E(X)\right)\left(Y - E(Y)\right)\right) \quad \dots (7.66)$$

Due to the similar reasons mentioned in (7.61), simplified form of (7.66) is

$$\text{Cov}(X, Y) = E\left(XY\right) - E\left(X\right)E\left(Y\right) \quad \dots (7.67)$$

Like expectation, we now discuss some properties of variance and covariance. Let us first state and prove some properties satisfied by variance operator/function V(.).

Statement: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and X, Y be random variables defined on the sample space Ω, then prove that

(a) $V(aX + b) = a^2\, V(X)$ ... (7.68)

(b) $V(aX) = a^2\, V(X)$ ... (7.69)

(c) $V(b) = 0$ ... (7.70)

(d) $V(aX \pm bY) = a^2\, V(X) + b^2\, V(Y) \pm 2ab\,\text{Cov}(X, Y)$ ... (7.71)

(e) $V(X \pm Y) = V(X) + V(Y) \pm 2\,\text{Cov}(X, Y)$ ... (7.72)

If X and Y are independent, then $\text{Cov}(X, Y) = 0$, so

(f) $V(aX \pm bY) = a^2\, V(X) + b^2\, V(Y)$ ... (7.73)

(g) $V(X \pm Y) = V(X) + V(Y)$ ... (7.74)

Proof: (a) By definition of variance refer (7.64), we have

$$\begin{aligned} V\left(aX + b\right) &= E\left[aX + b - E\left(ax + b\right)\right]^2 \\ &= E\left[aX + b - aE(X) - b\right]^2 \quad \left[\text{Using (7.20p1) to (7.20p3)}\right] \\ &= E\left[aX - aE(X)\right]^2 = E\left[a^2\left(X - E(X)\right)^2\right] \\ &= a^2 E\left[X - E(X)\right]^2 \quad \left[\text{Using (7.20p1)}\right] \\ &= a^2\, V(X) \quad \left[\text{Using (7.64)}\right] \quad \dots (7.75) \end{aligned}$$

(b) Putting b = 0 in (7.75), we get

$$V\left(aX + 0\right) = a^2\, V(X) \Rightarrow V\left(aX\right) = a^2\, V(X)$$

(c) Putting a = 0 in (7.75), we get

$$V\left(0X + b\right) = 0^2\, V(X) \Rightarrow V\left(b\right) = 0$$

(d) By definition of variance refer (7.64), we have

$$\begin{aligned} V\left(aX \pm bY\right) &= E\left[aX \pm bY - E\left(aX \pm bY\right)\right]^2 \\ &= E\left[aX \pm bY - aE(X) \mp bE(Y)\right]^2 \quad \left[\text{Using (7.20p1) to (7.20p3)}\right] \end{aligned}$$

$$= E\left[a\left(X - E(X)\right) \pm b\left(Y - E(Y)\right)\right]^2$$

Using the identity $(a \pm b)^2 = a^2 + b^2 \pm 2ab,$ we have

$$V\left(aX \pm bY\right) = E\begin{bmatrix} a^2\left(X - E(X)\right)^2 + b^2\left(Y - E(Y)\right)^2 \\ \pm 2ab\left(X - E(X)\right)\left(Y - E(Y)\right) \end{bmatrix}$$

$$= a^2 E\left(X - E(X)\right)^2 + b^2 E\left(Y - E(Y)\right)^2 \pm 2abE\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right]$$

$$\Rightarrow V\left(aX + bY\right) = a^2V(X) + b^2V(Y) \pm 2ab\,\mathrm{Cov}(X,\ Y) \qquad \ldots (7.76)$$

(e) Putting a = 1, b = 1 in (7.76), we have

$$V\left(1X \pm 1Y\right) = 1^2 V(X) + 1^2 V(Y) \pm 2(1)(1)\mathrm{Cov}(X,\ Y)$$

or $V\left(X \pm Y\right) = V(X) + V(Y) \pm \mathrm{Cov}(X,\ Y)$ … (7.77)

(f) Since random variables X and Y are independent, therefore

$$\mathrm{Cov}(X,\ Y) = 0 \qquad \ldots (7.78)$$

Using (7.78) in (7.76), we have

$$V\left(aX + bY\right) = a^2V(X) + b^2V(Y) \qquad \ldots (7.79)$$

(g) Using (7.78) in (7.77), we have

$$V\left(X \pm Y\right) = V(X) + V(Y)$$

Let us now state and prove some properties satisfied by covariance operator/function Cov(.).

Statement: Let $(\Omega,\ \mathcal{F},\ \mathcal{P})$ be a probability space and X, Y be random variables defined on the sample space $\Omega,$ then prove that

(a) $\mathrm{Cov}(X,\ a) = 0$ ... (7.80)

(b) $\mathrm{Cov}(aX,\ Y) = a\,\mathrm{Cov}\left(X,\ Y\right)$ ... (7.81)

(c) $\mathrm{Cov}(X,\ aY) = a\,\mathrm{Cov}\left(X,\ Y\right)$ ... (7.82a)

(d) $\mathrm{Cov}(aX,\ bY) = ab\,\mathrm{Cov}\left(X,\ Y\right)$ ... (7.82b)

(e) $\mathrm{Cov}(X,\ Y) = \mathrm{Cov}\left(Y,\ X\right)$ Symmetric property ... (7.83)

(f) $\mathrm{Cov}(X,\ X) = V\left(X\right)$ ... (7.84)

Bilinearity

(g) $\mathrm{Cov}(X + Y,\ Z) = \mathrm{Cov}\left(X,\ Z\right) + \mathrm{Cov}(Y,\ Z)$ ... (7.85)

(h) $\mathrm{Cov}(X,\ Y + Z) = \mathrm{Cov}\left(X,\ Y\right) + \mathrm{Cov}(X,\ Z)$ ... (7.86)

Proof: (a) By definition of covariance refer (7.66), we have

$$\begin{aligned} \mathrm{Cov}(X,\ a) &= E\left(\left(X - E(X)\right)\left(a - E(a)\right)\right) \\ &= E\left(\left(X - E(X)\right)\left(a - a\right)\right) && \left[\text{Using (7.20p4)}\right] \\ &= E(0) \\ &= 0 && \left[\text{Using (7.20p4)}\right] \end{aligned}$$

(b) By definition of covariance refer (7.66), we have

$$\begin{aligned}\mathrm{Cov}(aX,\ Y) &= E\big((aX - E(aX))(Y - E(Y))\big) \\ &= E\big((aX - aE(X))(Y - E(Y))\big) && [\text{Using (7.20p1)}] \\ &= E\big(a(X - E(X))(Y - E(Y))\big) \\ &= aE\big((X - E(X))(Y - E(Y))\big) && [\text{Using (7.20p1)}] \\ &= a\,\mathrm{Cov}(X,\ Y) && [\text{Using (7.66)}]\end{aligned}$$

Similarly, we can prove part (c).

Combining results of parts (b) and (c), you can easily prove part (d).

(e) By definition of covariance refer (7.66), we have

$$\begin{aligned}\mathrm{Cov}(X,\ Y) &= E\big((X - E(X))(Y - E(Y))\big) \\ &= E\big((Y - E(Y))(X - E(X))\big) && [\because E(UV) = E(VU)] \\ &= \mathrm{Cov}(Y,\ X) && [\text{Using (7.66)}]\end{aligned}$$

(f) By definition of covariance refer (7.66), we have

$$\begin{aligned}\mathrm{Cov}(X,\ X) &= E\big((X - E(X))(X - E(X))\big) \\ &= E\big(X - E(X)\big)^2 \\ &= V(X) && [\text{Using (7.64)}]\end{aligned}$$

(g) By definition of covariance refer (7.66), we have

$$\begin{aligned}\mathrm{Cov}(X + Y,\ Z) &= E\big[(X + Y - E(X + Y))(Z - E(Z))\big] \\ &= E\big[(X + Y - E(X) - E(Y))(Z - E(Z))\big] && [\text{Using (7.20p2)}] \\ &= E\big[(X - E(X) + Y - E(Y))(Z - E(Z))\big] \\ &= E\big[(X - E(X))(Z - E(Z)) + (Y - E(Y))(Z - E(Z))\big] \\ &= E\big[(X - E(X))(Z - E(Z))\big] + E\big[(Y - E(Y))(Z - E(Z))\big] && [\text{Using (7.20p2)}] \\ &= \mathrm{Cov}(X,\ Z) + \mathrm{Cov}(Y,\ Z) && [\text{Using (7.66)}]\end{aligned}$$

Combining (7.83) and (7.85), you can easily prove part (h).

Let us apply these properties in solving problems.

Example 5: Suppose in a bag there are 2 red, 4 blue and 5 black balls. Three balls are drawn from this bag randomly. Let X, Y denote the number of red and blue balls respectively, out of the three drawn balls. Find:

(a) E(X) (b) E(Y) (c) V(X) (d) V(Y) (e) Cov(X, Y) (f) SD(X) (g) Interpret results of parts (e) and (f).

Solution: We have already obtained joint PMF of (X, Y) refer Table 6.2 of Unit 6 of this course. In the same unit, the marginal probability mass functions of X and Y are shown in Table 6.3 as last column and last row respectively. So, for ready reference let us rewrite that table as Table 7.3 given as follows.

Table 7.3: Joint PMF of (X, Y) with marginal PMF's of X and Y respectively in last column and last row

	Values of Y				
Values of X	0	1	2	3	**Rows sums**
0	$\frac{10}{165}$	$\frac{40}{165}$	$\frac{30}{165}$	$\frac{4}{165}$	$\mathbf{\frac{84}{165}}$
1	$\frac{20}{165}$	$\frac{40}{165}$	$\frac{12}{165}$	0	$\mathbf{\frac{72}{165}}$
2	$\frac{5}{165}$	$\frac{4}{165}$	0	0	$\mathbf{\frac{9}{165}}$
Columns sums	$\mathbf{\frac{35}{165}}$	$\mathbf{\frac{84}{165}}$	$\mathbf{\frac{42}{165}}$	$\mathbf{\frac{4}{165}}$	$\mathbf{\frac{165}{165}} = 1$

Now, we can find required expected values, variances and covariance as follows.

(a) By definition of expected value refer (7.10), we have

$$E(X) = \sum_{x=0}^{2} x\,\mathcal{P}\left(X = x\right) = (0)\left(\frac{84}{165}\right) + (1)\left(\frac{72}{165}\right) + (2)\left(\frac{9}{165}\right)$$

$$= \frac{0+72+18}{165} = \frac{90}{165} = \frac{6}{11} \qquad \text{... (7.87)}$$

(b) Similarly, expected value of Y is given by

$$E(Y) = \sum_{y=0}^{3} y\,\mathcal{P}\left(Y = y\right) = (0)\left(\frac{35}{165}\right) + (1)\left(\frac{84}{165}\right) + (2)\left(\frac{42}{165}\right) + (3)\left(\frac{4}{165}\right)$$

$$= \frac{0+84+84+12}{165} = \frac{180}{165} = \frac{12}{11} \qquad \text{... (7.88)}$$

(c) Using (7.65) variance of random variable X is given by

$$\mu_X = V\left(X\right) = E\left(X^2\right) - \left(E(X)\right)^2$$

$$= \sum_{x=0}^{2} x^2\,\mathcal{P}\left(X = x\right) - \left(E(X)\right)^2$$

$$= (0)^2\left(\frac{84}{165}\right) + (1)^2\left(\frac{72}{165}\right) + (2)^2\left(\frac{9}{165}\right) - \left(\frac{6}{11}\right)^2$$

$$= \frac{0+72+36}{165} - \frac{36}{121} = \frac{1188-540}{165\times 11} = \frac{648}{165\times 11} = \frac{216}{605} \qquad \text{... (7.89)}$$

(d) Similarly, using (7.65) variance of random variable Y is given by

$$\mu_Y = V\left(Y\right) = E\left(Y^2\right) - \left(E(Y)\right)^2$$

$$= \sum_{x=0}^{3} y^2\,\mathcal{P}\left(Y = y\right) - \left(E(Y)\right)^2$$

$$= (0)^2\left(\frac{35}{165}\right) + (1)^2\left(\frac{84}{165}\right) + (2)^2\left(\frac{42}{165}\right) + (3)^2\left(\frac{4}{165}\right) - \left(\frac{12}{11}\right)^2$$

$$= \frac{0+84+168+36}{165} - \frac{144}{121} = \frac{288}{15\times 11} - \frac{144}{11\times 11} = \frac{3168-2160}{15\times 11\times 11}$$

$$= \frac{1008}{165\times 11} = \frac{336}{55\times 11} = \frac{336}{605} \qquad \text{... (7.90)}$$

(e) Using (7.65) covariance of random variables X and Y is given by

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$$
$$= \sum_{y=0}^{3}\sum_{x=0}^{2} x\,y\,\mathcal{P}(X = x, Y = y) - E(X)E(Y)$$
$$= (0)(0)\left(\frac{10}{165}\right) + (0)(1)\left(\frac{40}{165}\right) + (0)(2)\left(\frac{30}{165}\right) + (0)(3)\left(\frac{4}{165}\right) +$$
$$+ (1)(0)\left(\frac{20}{165}\right) + (1)(1)\left(\frac{40}{165}\right) + (1)(2)\left(\frac{12}{165}\right) + (1)(3)(0)$$
$$+ (2)(0)\left(\frac{5}{165}\right) + (2)(1)\left(\frac{4}{165}\right) + (2)(2)(0) + (2)(3)(0) - \frac{6}{11}\times\frac{12}{11}$$
$$= \frac{40+24+8}{165} - \frac{72}{121} = \frac{72}{15\times 11} - \frac{72}{11\times 11} = \frac{72(11-15)}{15\times 11\times 11}$$
$$= \frac{-288}{165\times 11} = \frac{-96}{55\times 11} = \frac{-96}{605} \qquad \ldots (7.91)$$

(f) We know that standard deviation of X is positive square root of variance of X. Hence, $SD(X) = \sqrt{\text{Variance of } X} = \sqrt{216/605} \simeq 0.5975.$

(g) **Interpretation of SD**: SD(X) can be interpreted as an average distance of values of X from expected value of X. So, average distance of values of X from expected value 6/11 of X is 0.5975 (approx.).

Interpretation of Covariance: From (7.91) we see that covariance is negative which means that if X increases/decreases then Y will decrease/increase. In the case when covariance is positive then it means X and Y increase or decrease together. That is, if X increases then Y will also increase and if X decreases then Y will also decrease.

At the end of the first paragraph of this section we promised that in the end of this section, we will discuss some important properties of MGF. So, it is time to keep that promise.

Properties of MGF: State and prove properties of MGF.

Statement: If X is a random variable discrete or continuous then prove that MGF satisfies the following properties.

(a) **Translation Property**: $M_{cX}(t) = M_X(ct), \text{ where } c \text{ is a real number}$ .. (7.92)

(b) **Change of Origin and Scale Property**: … (7.93)

$$\text{If } Y = \frac{X-a}{h}, \text{ then } M_Y(t) = e^{-\frac{a}{h}t} M_X\left(\frac{t}{h}\right), \text{ where } a, h \text{ are real numbers}$$

(c) **Sum to Product Property when Random Variables are Independent**: If random variables X and Y are independent, then $M_{X+Y}(t) = M_X(t)M_Y(t)$

In general if $X_1, X_2, X_3, \ldots, X_n$ are independent random variables then

$$M_{X_1+X_2+X_3+\ldots+X_n}(t) = M_{X_1}(t)M_{X_2}(t)M_{X_3}(t)\ldots M_{X_n}(t) \qquad \ldots (7.94)$$

(d) **Uniqueness Property**: If $M_X(t) = M_Y(t)$, then random variables X and Y have the same probability distribution. … (7.95)

Proof: By definition of MGF (you may refer to 7.48), $M_X(t) = E\left(e^{tX}\right) \ldots$ (7.96)

(a) Using (7.96), we have $M_{cX}(t) = E\left(e^{t(cX)}\right) = E\left(e^{(ct)X}\right) = M_X(ct).$

(b) Using (7.96), we get

$$M_Y(t) = E\left(e^{tY}\right) = E\left(e^{t\left(\frac{X-a}{h}\right)}\right) = E\left(e^{\left(\frac{tX}{h}-\frac{at}{h}\right)}\right) = E\left(e^{-\frac{at}{h}} e^{\frac{tX}{h}}\right)$$

$$= e^{-\frac{at}{h}} E\left(e^{\frac{tX}{h}}\right) \quad \left[\begin{array}{l}\because E(aX) = aE(X), \text{ where a is constant.} \\ \text{You may refer to (7.20p1). Here } a = e^{-\frac{at}{h}}\end{array}\right]$$

$$= e^{-\frac{at}{h}} E\left(e^{\left(\frac{t}{h}\right)X}\right)$$

$$= e^{-\frac{a}{h}t} M_X\left(\frac{t}{h}\right), \text{ where } a, h \text{ and } t \text{ are real numbers}$$

(c) **Sum to Product Property when Random Variables are Independent**:
Using (7.96), we get

$$M_{X+Y}(t) = E\left(e^{t(X+Y)}\right) = E\left(e^{(tX+tY)}\right) = E\left(e^{tX} e^{tY}\right)$$

$$= E\left(e^{tX}\right)E\left(e^{tY}\right) \left[\begin{array}{l}\because \text{If } X \text{ and } Y \text{ are independent random variables,} \\ \text{then so will be } e^{tX} \text{ and } e^{tY} \text{ and hence using} \\ \text{multiplication theorem of expectation} \\ E(XY) = E(X)E(Y)\end{array}\right]$$

$$= M_X(t) M_Y(t), \text{ where } t \text{ is real number}$$

Similarly, we can prove that

$$M_{X_1+X_2+X_3+\ldots+X_n}(t) = M_{X_1}(t) M_{X_2}(t) M_{X_3}(t) \ldots M_{X_n}(t)$$

(d) **Uniqueness Property**: Proof is beyond the scope of the course.

Example 6: If MGF of the random variable X is $\frac{1}{\sqrt{1-2t}}, t < \frac{1}{2}.$ Find MGF of Y = 5X.

Solution: We are given that

$$M_X(t) = \frac{1}{\sqrt{1-2t}}, t < \frac{1}{2} \qquad \ldots (7.97)$$

$$\therefore M_Y(t) = M_{5X}(t) = \frac{1}{\sqrt{1-2(5t)}} = \frac{1}{\sqrt{1-10t}}, 5t < \frac{1}{2} \quad [\text{Using } (7.92)]$$

$$\text{Or } M_Y(t) = M_{5X}(t) = \frac{1}{\sqrt{1-10t}}, t < \frac{1}{10}.$$

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

SAQ 2

If random variables X and Y are as defined in Example 5, then find the value of V(2X + 3Y).

7.4 SUMMARY

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

- In probability theory, weighted mean is known as the **expected value** where weights are probabilities. For discrete random variable it is defined by $E(X)=\sum_{x} x\,\mathcal{P}(X=x)$ and for continuous random variable it is defined by $E(X)=\int_{-\infty}^{\infty} x\,f_X(x)\,dx$.
- Some properties of expectation are given as follows.
  - E(aX) = a E(X)
  - E(X + Y) = E(X) + E(Y)
  - In general, $E(X_1+X_2+X_3+\ldots+X_n)=E(X_1)+E(X_2)+E(X_3)+\ldots+E(X_n)$
  - E(a) = a, where a is a constant
  - If $X(\omega)\geq 0 \quad \forall \ \omega\in\Omega,$ then E(X) ≥ 0
  - If $X(\omega)\geq Y(\omega) \quad \forall \ \omega\in\Omega,$ then E(X) ≥ E(Y)
- $|E(X)|\leq E(|X|)$ for any random variable X, provided expectations exist
- **Addition Theorem of Expectation**: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and X, Y be two random variables defined on the sample space Ω then addition theorem of expectation says that E(X + Y) = E(X) + E(Y).

  In general, if we have n random variables $X_1, X_2, X_3, \ldots, X_n$ defined on the sample space Ω then
  $E(X_1+X_2+X_3+\ldots+X_n)=E(X_1)+E(X_2)+E(X_3)+\ldots+E(X_n)$

  Further, if we have a countable sequence of random variables $X_1, X_2, X_3, \ldots$ defined on the sample space Ω then
  $E(X_1+X_2+X_3+\ldots)=E(X_1)+E(X_2)+E(X_3)+\ldots$
- **Multiplication Theorem of Expectation**: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and X, Y be two independent random variables defined on the sample space Ω then multiplication theorem of expectation says that

  E(XY) = E(X)E(Y).
- **r^th raw moment** is denoted by μ'_r and is defined by $\mu'_r=E(X^r)$.
- **r^th central moment** is denoted by μ_r and is defined by

  $\mu_r=E(X-\mu)^r$ or $\mu_r=E(X-E(X))^r$
- **r^th standardised moment** if denoted by μ_r^{sd} is defined by

  $\mu_r^{sd}=E\left(\frac{X-\mu}{\sigma}\right)^r$ or $\mu_r^{sd}=E\left(\frac{X-E(X)}{\sigma}\right)^r$
- $\sigma=\text{Standard deviation}=\sqrt{E(X-E(X))^2}$

- **Moment Generation Function (MGF)**: Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and X be a random variable defined on the sample space Ω. Let $A = \{t \in \mathbb{R} : E(e^{tX}) \text{ exists}\}$ then the function $M_X : A \to \mathbb{R}$ defined by

  $M_X(t) = E(e^{tX}),\ t \in A$ is called MGF of random variable X.

  Moments from MGF

  $M_X^{(1)}(0) = \mu'_1 \qquad M_X^{(2)}(0) = \mu'_2 \qquad M_X^{(3)}(0) = \mu'_3 \quad \ldots \quad M_X^{(r)}(0) = \mu'_r \quad \ldots$

  $\text{Variance of the distribution} = \mu_2 = E(X - E(X))^2 = E(X^2) - (E(X))^2$

  Covariance of X and Y it is defined by

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

- **Properties of variance are listed as follows**:
  - $V(aX + b) = a^2 V(X)$
  - $V(aX) = a^2 V(X)$
  - $V(b) = 0$
  - $V(aX \pm bY) = a^2 V(X) + b^2 V(Y) \pm 2ab\,\text{Cov}(X, Y)$
  - $V(X \pm Y) = V(X) + V(Y) \pm 2\,\text{Cov}(X, Y)$
  - If X and Y are independent, then $\text{Cov}(X, Y) = 0$, so

    $V(aX \pm bY) = a^2 V(X) + b^2 V(Y)$
    $V(X \pm Y) = V(X) + V(Y)$

- **Properties of covariance are listed as follows**:
  - $\text{Cov}(X, a) = 0$
  - $\text{Cov}(aX, Y) = a\,\text{Cov}(X, Y)$
  - $\text{Cov}(X, aY) = a\,\text{Cov}(X, Y)$
  - $\text{Cov}(aX, bY) = ab\,\text{Cov}(X, Y)$
  - $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
  - $\text{Cov}(X, X) = V(X)$
  - Bilinearity

    $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
    $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$

7.5 TERMINAL QUESTIONS

1. **Does expected value exist**? Recall our two old friends Anjali and Prabhat. Prabhat is a good statistician as well as a nice human being who always ready to help anyone other than near and dear ones also. One day they visit to a fair and his friend Anjali shows her interest in playing a game of coin toss. Rules of the game are: If Anjali gets the first head in an

odd trial then she will get Rs $\frac{2^n}{n}$, but if Anjali gets the first head in an even trial then she has to pay Rs $\frac{2^n}{n}$. Till the time his friend Anjali plays the game, Prabhat being a good statistician do some calculation and observe that there is a problem with long run value of this game because it depends on the order of the outcomes. Further, he observes that if terms are adjusted in a particular order, then long run value of the game will be log 2. You being a student of probability theory can explain the calculation used by Prabhat to get the argument behind his observations. Explain each step of the calculation done by Prabhat. Assume that coin used in the game is unbiassed/fair.

2. If random variables X and Y are independent and V(X) = 5, V(Y) = 3 then find V(X + Y) and V(X – Y).

3. In the statement of multiplication theorem of expectation there was one restriction that both random variables should be independent. Give an example, in which this restriction does not hold and hence conclusion of the theorem fails to hold.

4. If MGF of a random variable X is $\exp\left(\lambda\left(e^t - 1\right)\right)$ then find MGF of the random variable $Y = \frac{X-5}{10}$. For those learners who are not aware of the notation exp(), note that the notation exp(a) means e^a. Recall that in R, we use this notation for exponential function having base e, i.e., $e^5 = \exp(5)$.

5. We know that sum of deviations about expected value is always zero. So, the first central moment (μ_1) is always zero. From (7.63), you know that

$$\mu_2 = \mu_2' - \left(\mu_1'\right)^2. \qquad \ldots (7.98)$$

Obtain expression for third and fourth central moment in terms of first four raw moments of the distribution.

7.6 SOLUTIONS/ANSWERS

Self-Assessment Questions (SAQs)

1. We are given that random variable X has following PMF

$$\mathcal{P}\left(X = n\right) = \frac{1}{2^{n+1}}, \; n = 0, 1, 2, 3, \ldots$$

Using definition of expected value (refer 7.11) for discrete random variable, we have

$$E(X) = \sum_{n=0}^{\infty} n\,\mathcal{P}\left(X = n\right)$$

$$= \sum_{n=0}^{\infty} n\,\frac{1}{2^{n+1}} = (0)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{4}\right) + (2)\left(\frac{1}{8}\right) + (3)\left(\frac{1}{16}\right) + (4)\left(\frac{1}{32}\right) + \ldots$$

$$= \frac{1}{4}\left[1 + 2\left(\frac{1}{2}\right) + 3\left(\frac{1}{4}\right) + 4\left(\frac{1}{8}\right) + \ldots\right]$$

$$= \frac{1}{4}\left[1 + 2\left(\frac{1}{2}\right) + 3\left(\frac{1}{2}\right)^2 + 4\left(\frac{1}{2}\right)^3 + \ldots\right] \qquad \ldots (7.99)$$

We know that if $|x| < 1$, then sum of the infinite GP having common ratio x is given by

$$1 + x + x^2 + x^3 + x^4 + \ldots = \frac{1}{1-x} \qquad \ldots (7.100)$$

Differentiating w. r. t. x, we get

$$1 + 2x + 3x^2 + 4x^3 + \ldots = \frac{1}{(1-x)^2} \qquad \ldots (7.101)$$

Using (7.101) in (7.99), we get

$$E(X) = \frac{1}{4}\left[\frac{1}{(1-1/2)^2}\right] = \frac{1}{4}(4) = 1$$

2. Using (7.71), we have

$$V(2X + 3Y) = 4V(X) + 9V(Y) + 12\,Cov(X, Y)$$

$$= 4\left(\frac{216}{605}\right) + 9\left(\frac{336}{605}\right) + 12\left(\frac{-96}{605}\right) = \frac{864 + 3024 - 1152}{605} = \frac{2736}{605} \qquad \ldots (7.102)$$

Terminal Questions

1. Let random variable X denote gain of Anjali in this game then as per rules of the game, we have

$$X = \begin{cases} \dfrac{2^n}{n}, & \text{if } n = 1, 3, 5, 7, \ldots \\ -\dfrac{2^n}{n}, & \text{if } n = 2, 4, 6, 8, \ldots \end{cases}$$

To know long run behaviour of the game, we have to find out expected value of X. But before that we have to obtain probabilities of appearing the first head in odd number of trials as well as in even number of trials. Since coin is fair so,

$$\mathcal{P}(\text{getting head in any trial}) = \mathcal{P}(\text{getting tail in any trial}) = \frac{1}{2} \qquad \ldots (7.103)$$

Using (7.103) repeatedly and independence of trials, we have

$$\mathcal{P}(\text{getting first head on } n^{th} \text{ trial}) = \mathcal{P}\begin{pmatrix}\text{getting tail till} \\ n-1 \text{ trials}\end{pmatrix} \times \mathcal{P}\begin{pmatrix}\text{getting head} \\ \text{on } n^{th} \text{ trial}\end{pmatrix}$$

$$= \underbrace{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \ldots \frac{1}{2}}_{n-1 \text{ times}} \times \left(\frac{1}{2}\right) = \frac{1}{2^n}, \; n = 1, 2, 3, \ldots$$

Using definition of expected value (refer 7.11) for discrete random variable, we have

$$E(X) = \sum_{n=1}^{\infty} n\mathcal{P}(X = n) = \sum_{k=1}^{\infty} (2k-1)\mathcal{P}(X = 2k-1) + \sum_{k=1}^{\infty} (2k)\mathcal{P}(X = 2k)$$

$$= \sum_{k=1}^{\infty} \frac{2^{2k-1}}{2k-1}\frac{1}{2^{2k-1}} + \sum_{k=1}^{\infty} -\frac{2^{2k}}{2k}\frac{1}{2^{2k}} \qquad [\text{Using } (7.103)]$$

$$= \sum_{k=1}^{\infty} \frac{1}{2k-1} - \sum_{k=1}^{\infty} \frac{1}{2k} = \left(\frac{1}{1} + \frac{1}{3} + \frac{1}{5} + \ldots\right) - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \ldots\right)$$

$$= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \ldots \qquad \left[\begin{array}{l}\text{Adjusting terms by taking one} \\ \text{term at a time from first and} \\ \text{second brackets alternatively}\end{array}\right]$$

$$= \log(1+1) \qquad \left[\begin{array}{l}\because \log(1+x) = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \dfrac{x^4}{4} + \ldots \\ \text{In our case } x = 1.\end{array}\right]$$

$$= \log 2$$

Refer Example 15 of Unit 4 of the course MST-011, this infinite alternating series converges conditionally not absolutely. So, expected value of this game does not exist due to definition of the expected value, refer (7.7), (7.8) and (7.9) of this unit. This completes the explanation behind the calculation done by our old friend Prabhat.

2. We know that if random variables X and Y are independent then $V(X \pm Y) = V(X) + V(Y)$

$\therefore\ V(X+Y) = V(X) + V(Y) = 5 + 3 = 8$ and also
$V(X-Y) = V(X) + V(Y) = 5 + 3 = 8.$

3. Consider a random experiment of tossing a fair coin twice. Sample space of this random experiment is $\Omega = \{HH, HT, TH, TT\}$. Let X be the random variable which counts the number of heads and Y be another random variable which counts the number of tails. So, we have

$X(HH) = 2,\ X(HT) = 1,\ X(TH) = 1,\ X(TT) = 0,$ and
$Y(HH) = 0,\ Y(HT) = 1,\ Y(TH) = 1,\ Y(TT) = 2.$

Note that information about one random variable tells either complete or partial information about the other random variable. For example, if X = 2 then it tells complete information about Y that none of the two outcomes is tail and so Y = 0. If X = 1 then it tells partial information about Y that there is one tail so Y = 1 but we do not know whether the first outcome is tail or the second outcome is tail. Since one random variable tells complete or partial information about the other so the two random variables X and Y are not independent. In fact, Y = 2 – X. Hence, the restriction of independence in the statement of multiplication theorem of expectation does not hold. Now, we will show that conclusion of the theorem also does not hold. To show it, we have to obtain three expected values E(X), E(Y) and E(XY) separately. But before that we have to write joint PMF of discrete random variable (X, Y) together with sums of rows and columns of probabilities to get marginal PMF's of X and Y. Table 7.4 shows joint and marginal PMF's.

Table 7.4: Joint PMF of (X, Y) with marginal PMF's of X and Y respectively in last column and last row

	Values of Y			
Values of X	0	1	2	**Rows sums**
0	0	0	1/4	**1/4**
1	0	2/4	0	**2/4**
2	1/4	0	0	**1/4**
Columns sums	**1/4**	**2/4**	**1/4**	1

By definition of expected refer (7.10), we have

$$E(X) = \sum_{x=0}^{2} x\,\mathcal{P}\left(X = x\right) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{2}{4}\right) + (2)\left(\frac{1}{4}\right) = \frac{2+2}{4} = \frac{4}{4} = 1. \quad (7.104)$$

$$E(Y) = \sum_{x=0}^{2} y\,\mathcal{P}\left(Y = y\right) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{2}{4}\right) + (2)\left(\frac{1}{4}\right) = \frac{2+2}{4} = \frac{4}{4} = 1. \quad (7.105)$$

$$E(XY) = \sum_{y=0}^{2}\sum_{x=0}^{2} x\,y\,\mathcal{P}\left(X = x\right)\mathcal{P}\left(Y = y\right)$$

$$= (0)(0)\left(0\right) + (0)(1)\left(0\right) + (0)(2)\left(\frac{1}{4}\right) + (1)(0)\left(0\right) + (1)(1)\left(\frac{2}{4}\right) + (1)(2)\left(0\right)$$

$$+ (2)(0)\left(\frac{1}{4}\right) + (2)(1)\left(0\right) + (2)(2)\left(0\right) = \frac{2}{4} = \frac{1}{2} \quad \ldots (7.106)$$

From (7.104) and (7.105), we have $E(X)E(Y) = (1)(1) = 1$ … (7.107)

From (7.106) and (7.107), we have $E(X\,Y) \neq E(X)E(Y)$.

Hence conclusion of multiplication theorem of expectation does not hold.

4. We are given that $M_X(t) = \exp\left(\lambda\left(e^t - 1\right)\right)$. … (7.108)

We know that MGF of $Y = \dfrac{X-5}{10}$ in terms of MFG of X is given by (you may refer 7.93)

$$M_Y(t) = e^{-\frac{a}{h}t} M_X\left(\frac{t}{h}\right)$$

$$\therefore\ M_{\frac{X-5}{10}}(t) = e^{-\frac{5}{10}t} M_X\left(\frac{t}{10}\right)$$

$$= e^{-\frac{5}{10}t} \exp\left(\lambda\left(e^{\frac{t}{10}} - 1\right)\right) \quad \left[\text{Using } (7.108)\right]$$

5. Using (7.45) 3rd central moment is given by

$$\mu_3 = E\left(X - E(X)\right)^3$$

$$= E\left(\begin{aligned}&\binom{3}{0} X^3 \left(E(X)\right)^0 - \binom{3}{1} X^{3-1} \left(E(X)\right)^1 + \binom{3}{2} X^{3-2} \left(E(X)\right)^2 \\ &- \binom{3}{3} X^{3-3} \left(E(X)\right)^3\end{aligned}\right)$$

$$\left[\text{Using binomial expansion for } n = 3\right]$$

$$= E\left(X^3 - 3X^2 E(X) + 3X(E(X))^2 - (E(X))^3\right)$$

$$= E\left(X^3 - 3\mu_1' X^2 + 3(\mu_1')^2 X - (\mu_1')^3\right) \qquad \left[\text{Using (7.44)}\right]$$

$$= E\left(X^3\right) - 3\mu_1' E(X^2) + 3(\mu_1')^2 E(X) - (\mu_1')^3$$

$$= \mu_3' - 3\mu_1'\mu_2' + 3(\mu_1')^2 \mu_1' - (\mu_1')^3 \qquad \left[\text{Using (7.44)}\right]$$

$$= \mu_3' - 3\mu_1'\mu_2' + 3(\mu_1')^3 - (\mu_1')^3$$

$$= \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3$$

$$\Rightarrow \mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3 \qquad \ldots (7.109)$$

Similarly, 4th central moment using (7.45) is given by

$$\mu_4 = E\left(X - E(X)\right)^4$$

$$= E\begin{pmatrix} \binom{4}{0} X^4 (E(X))^0 - \binom{4}{1} X^{4-1} (E(X))^1 + \binom{4}{2} X^{4-2} (E(X))^2 \\ -\binom{4}{3} X^{4-3} (E(X))^3 + \binom{4}{4} X^{4-4} (E(X))^4 \end{pmatrix}$$

$$\left[\text{Using binomial expansion for } n = 4\right]$$

$$= E\left(X^4 - 4X^3 E(X) + 6X^2 (E(X))^2 - 4X(E(X))^3 + (E(X))^4\right)$$

$$= E\left(X^4 - 4\mu_1' X^3 + 6(\mu_1')^2 X^2 - 4X(\mu_1')^3 + (\mu_1')^4\right) \qquad \left[\text{Using (7.44)}\right]$$

$$= E\left(X^4\right) - 4\mu_1' E(X^3) + 6(\mu_1')^2 E(X^2) - 4(\mu_1')^3 E(X) + (\mu_1')^4$$

$$= \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2 \mu_2' - 4(\mu_1')^3 \mu_1' + (\mu_1')^4 \qquad \left[\text{Using (7.44)}\right]$$

$$= \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2 \mu_2' - 4(\mu_1')^4 + (\mu_1')^4$$

$$= \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2 \mu_2' - 3(\mu_1')^4$$

$$\Rightarrow \mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2 \mu_2' - 3(\mu_1')^4 \qquad \ldots (7.110)$$