2
CHAPTER

Random Variables

2.1 INTRODUCTION

Many random phenomena in engineering are associated with numerical outcomes of some physical quantity. The possible outcome of a random phenomenon can be identified numerically or artificially. *A random variable is a numerical variable whose specific value cannot be predicted with certainty before an experiment*. The outcomes or events are identified by values of a function and such a function is called a random variable. Hence, a random variable is a function of sample space of an experiment such that there is a numerical value of the random variable corresponding to each outcome of an experiment. Different simple events in engineering like occurrence of earthquakes, strength of concrete, traffic count, runoff in hydrological studies, etc., may have some associated values relating to its certainty/uncertainty. The value assumed by a random variable depends on the outcome of the experiment and has probability associated with each occurrence of event in the sample space. In short, random variable is a tool or device to identify events in numerical terms. The probability measures associated with a numerical value of a random variable is described by probability distribution or probability law, i.e. Probability Mass Function (PMF)/Probability Density Function (PDF) and Cumulative Density Function (CDF).

2.2 TYPES OF RANDOM VARIABLE

Random variables can be discrete or continuous. Random variable can assume either integer values or fractional values. A variable which assumes integer values is called discrete variable and which assumes fractional value is called continuous variable.

2.2.1 Characteristic of a Random Variable

In science and engineering, we are interested in numerical outcomes. There are two types of outcome in random experiment, namely, numerical and descriptive. In a dice experiment we obtain the numerical (outcomes as 1, 2, 3, 4, 5, and 6) while tossing a coin the outcome turnout as head or tail, hence the sample space in non-numerical (outcome as head or tail) or descriptive. Instead of dealing with descriptive values, it is easy to assign numerical value say, ‘1’ to head and ‘0’ to tail. The interpretation is easy and attractive from mathematical point of view. This rule of *mapping* from the original sample

(b) When $x = 0, 1, 2, 3, 4$ for the function $p_X(x) = x^2/25$, we have

$$p_X(0) = 0,\ p_X(1) = 1/25,\ p_X(2) = 4/25,\ p_X(3) = 9/25 \text{ and } p_X(4) = 16/25.$$

The value of the function, that is, the probabilistic values for each value of x from 0 to 4 are equal to or greater than zero and Eq. (2.1) is satisfied. But the summation of all the probabilistic value of all events is

$$\sum_{i=0}^{4} p_X(x_i) = 0 + \frac{1}{25} + \frac{4}{25} + \frac{9}{25} + \frac{16}{25}$$

$$= \frac{30}{25} \text{ and greater than one, which violates Eq. (2.2).}$$

Hence, $p_X(x) = x^2/25$ for $x = 0, 1, 2, 3, 4$ is not a valid function to serve as PMF.

Example 2.2 A contractor is planning to purchase 3 bulldozers for a new project in a remote area. The possible condition of bulldozers after 6 months is either good (G), or bad (B). What are the possible statuses of the bulldozers?

Solution: A random experiment consists of independent toss of a fair single coin (bulldozer) and $H(G)$ denotes head and $T(B)$ denotes tail. If three coins (3 bulldozers) are tossed the possible statuses (sample space) is shown below in the probability table of random variable X.

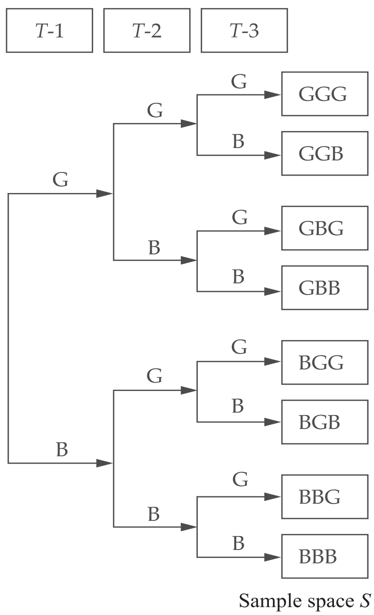

Fig. 2.2 Tree diagram

S = (GGG, GBB, BGB, BBG, GGB, GBG, BGG, BBB) the probability of each outcome S is 1/8.

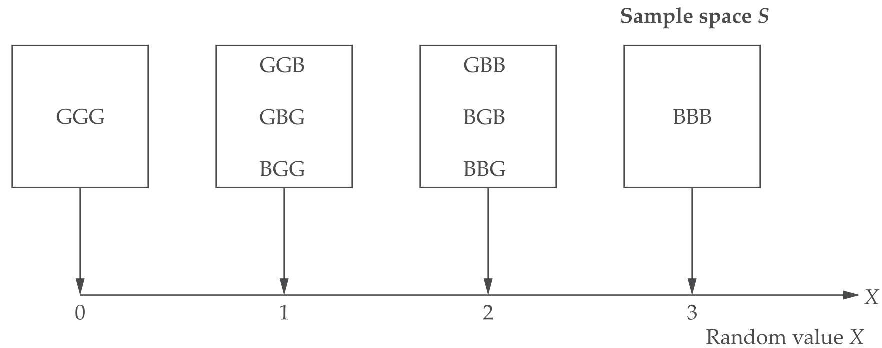

Fig. 2.3 Mapping of sample space on random value

Probability table of random variable *X*

X	0	1	2	3
$p_X(x)$	1/8	3/8	3/8	1/8

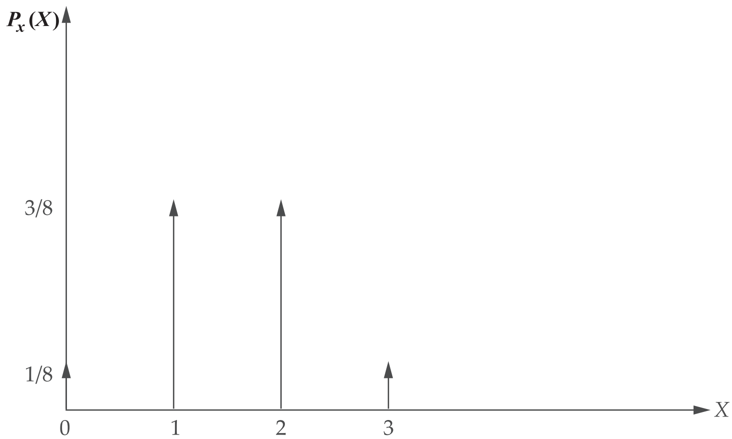

Fig. 2.4 Probability mass function

The Cumulative Distribution Function (CDF) of X can easily be written from the above table

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ (1/8) & \text{if} \quad 0 \le x < 1 \\ (4/8) & \text{if} \quad 1 \le x < 2 \\ (7/8) & \text{if} \quad 2 \le x < 3 \\ 1 & \text{if} \quad x \ge 3 \end{cases}$$

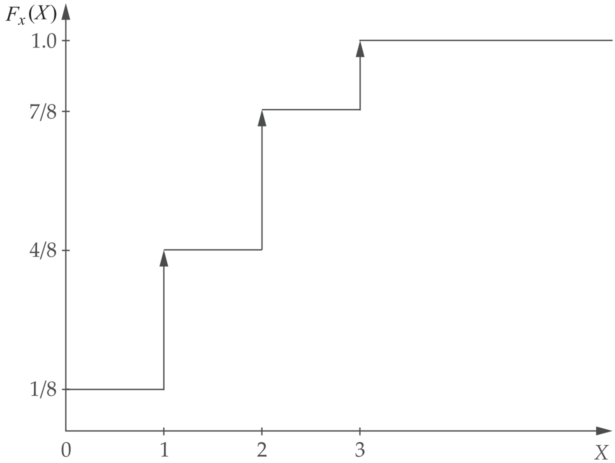

Fig. 2.5 Cumulative distribution function

2.3 TWO-DIMENSIONAL RANDOM VARIABLES

A two-dimensional random variable is mapping of the points in the sample space to ordered pairs $[x, y]$. Usually, when dealing with a pair of random variables, the sample space naturally partitions itself so that it can be viewed as a combination of two simpler sample spaces. For example, if the experiment was to observe the density and weight of soil. The range of soil density could fall in some set, which we will call sample space S_1, while the range of the weight would fall within sample space S_2. The overall sample space of the experiment could be viewed as $S = (S_1 \times S_2)$. The outcome of the experiment, the pair of random variables (X, Y) is merely a mapping of the outcome S to a pair of numerical values $(x(S), y(S))$. It would be natural to choose $x(S)$ to be the density (kN/m^3) while $y(S)$ is the weight of soil in (kN). The density functions of $f_X(x)$ and $f_Y(y)$ do partially characterize the experiment; they do not completely describe the situation. We accomplish this through the joint cumulative distribution function and the joint density function.

2.4 JOINT PROBABILITY MASS FUNCTION

When the random variables are discrete, it is convenient to work with probability mass function straightforward and it is possible to extend the concept to a pair of random variables X, Y given by $P_{X,Y}(x, y) = P\,(\{X = x\} \cap \{Y = y\})$. In particular, suppose the random variable X takes on values from the set $\{x_1, x_2, ...\, x_M\}$ and the random variable Y takes on values from the set $\{y_1, y_2, y_3 \dots y_N\}$. Here either M and/or N could be potentially infinite or both could be finite.

Properties of joint probability mass function:

(1) $0 \le P_{X,Y}\left(x_m, y_n\right) \le 1$ …(2.4)

(2) $\sum_{m=1}^{M}\sum_{n=1}^{N} P_{X,Y}\left(x_m, y_n\right) = 1$ …(2.5)

(3) $\sum_{n=1}^{N} P_{X,Y}(x,y) = P_X(x_m)$ ...(2.6)

(4) $\sum_{m=1}^{M} P_{X,Y}(x,y) = P_Y(y_n)$ ...(2.7)

Further, the joint PDF or the joint CDF of a pair of discrete random variables can be related to the joint PMF through the use of delta function or step function by

(5) $f_{X,Y}(x,y) = \sum_{m=1}^{M}\sum_{n=1}^{N} P_{X,Y}(x_m, y_n)\delta(x-x_m)\delta(y-y_n)$ ...(2.8)

(6) $f_{X,Y}(x,y)\sum_{m}^{M}\sum_{n}^{N} P_{X,Y}(x_m, y_n)\delta(x-x_m)\delta(y-y_n)$ ...(2.9)

The PMF of random variable X given for the value of Y, then, by the definition of conditional probability is:

(7) $P(X=x|Y=y) = \dfrac{P(X=x, Y=y)}{P(Y=y)} = \dfrac{P_{X,Y}(x,y)}{P_Y(y)}$ ...(2.10)

We refer to this as the conditional PMF of X given Y. By way of notation we write

$$P_{X|y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)} \qquad ...(2.11)$$

2.5 MATHEMATICAL EXPECTATION

The mean of a random variable is also called its expected value. The notation $E(X)$ is used to denote the expected value. Thus, the mean, μ_X and expected value $E(X)$ are the same quantity and can be used interchangeably.

The mean of a discrete random variable X or the population mean is expressed as

$$E(X) = \Sigma(\text{value}) \times (\text{probability}) = \sum_{\text{all}\cdot x_i} x_i\, p_X(x_i) \qquad ...(2.12)$$

Here, the sum extends over all the distinct values of X.

The variance, Var(X) is given by the expression

$$\text{Var}(X) = \sigma_X^2 = \sum_{\text{all}\cdot x_i} (x_i - \mu)^2 \times p_X(x_i) \qquad ...(2.13)$$

i.e., summation of the product of squares of deviation from mean and the corresponding probability.

In general the r^{th} moment about the mean, μ is given by $E(X-\mu)^r$ and the expectation of $E(X^r)$ is given by $\Sigma(x_i - \mu)^r \times p_X(x_i)$.

Suppose x is a discrete random variable which takes the values $x_1, x_2, x_3, \ldots\ldots. x_n$ with respect to probabilities $p_X(x_1), p_X(x_2), p_X(x_3), \ldots\ldots. p_X(x_n)$ then it is well known that from Eq. (2.2), $\Sigma p_X(x_i)$ equals unity or

$$p_X(x_1)+p_X(x_2)+p_X(x_3)+........+p_X(x_n)=1.00$$

The mathematical expectation of random variable, X or the expected value of X can be obtained by treating probabilities $p_X(x_1), p_X(x_2), p_X(x_3), \ldots\ldots. p_X(x_n)$ as loads acting at a distance $x_1, x_2, x_3, \ldots\ldots. x_n$ from a point. The resultant of the force system is then given by

$$\mu=\frac{p_X(x_1)x_1+p_X(x_2)x_2+......+p_X(x_n)x_n}{p_X(x_1)+p_X(x_2)+......+p_X(x_n)} \quad \ldots(2.14)$$

or $$E(X) = \mu_X=\frac{\sum_{\text{all}\cdot x_i} p_X(x_i)\times x_i}{\sum_{\text{all}\cdot x_i} p_X(x_i)} \quad \ldots(2.15)$$

Since $\sum_{\text{all}\cdot x_i} p_X(x_i) = 1,$

$$E(X) = \mu_X \quad \ldots(2.16)$$

As stated earlier that variance of X is given by $E(X - \mu_X)^2$

Therefore $$\text{Var}(X),\ \sigma_x^2 = E(X - \mu_X)^2 \quad \ldots(2.17)$$

or $$\sigma_x^2 = E(X^2+\mu_X^2-2X\mu_X)$$

$$= E(X^2)+\mu_X^2-2\mu_X E(X)$$

$$= E(X^2)+\mu_X^2-2\mu_X^2$$

$$= E(X^2) - \mu_X^2$$

Hence, $$E(X^2)=\sigma_X^2+\mu_X^2 \quad \ldots(2.18)$$

The third moment of X about the mean gives the skewness

i.e. $$\text{Skewness} = E(X - \mu_X)^3 \quad \ldots(2.19)$$

The skewness coefficient, r_1 is given by $\dfrac{E(X-\mu_X)^3}{\sigma_X^3}$ …(2.20)

The fourth moment of X about the mean gives the Kurtosis

i.e. $$\text{Kurtosis} = E(X - \mu_X)^4 \quad \ldots(2.21)$$

The coefficient of kurtosis, r_2 is given by $\dfrac{E(X-\mu_X)^4}{\sigma_X^4}$ …(2.22)

Mean of the discrete random variable, $X = \mu_X$, is given

$$= E(X) = \sum_{\text{all}\cdot x_i} x_i \times p_X(x_i) = \frac{252}{36}$$

$$= 7$$

$$\text{Var }(X) = \sigma_X^2 = \sum_{\text{all}\cdot x_i} (x-\mu)^2\, p_X(x_i) = \frac{210}{36}$$

$$= 5.833$$

Standard deviation of variable, $\sigma_X = \sqrt{\text{Var}(X)}$

$$= 2.415$$

Coefficient of variation, $(\text{CoV})_X = \dfrac{\sigma_X}{\mu_X}$

$$= \frac{2.415}{7}$$

$$= 0.345$$

Coefficient of skewness, $r_1 = \dfrac{E(X-\mu_X)^3}{\sigma_X^3}$

$$= \frac{\sum_{\text{all}\cdot x} (x-\mu_X)^3 \times p_X(x)}{\sigma_X^3}$$

= 0, hence, Symmetrical Distribution

Coefficient of kurtosis, $r_2 = \dfrac{E(X-\mu_X)^4}{\sigma_X^4}$

$$= \frac{\sum_{\text{all}\cdot x_i} (x-\mu_X)^4 \times p_X(x_i)}{\sigma_X^4} = \left(\left(\frac{2898}{36}\right) \Big/ \left(\frac{210}{36}\right)^2 \right)$$

$$= 2.36$$

2.6 ALGEBRA OF EXPECTATION

The expectations possess certain properties. For a random variable, expectation of a constant is a constant.

Hence,

$$E(a) = ax \qquad \text{… (2.23)}$$

$$E(aX) = a \times E(X)$$

$$= a\mu_X \qquad \text{…(2.24)}$$

where μ_X is the mean of the variable X

Supposing $$Y = a + bX$$

Then
$$\begin{aligned} E[Y] &= E[a + bX] \\ &= a + bE[X] \\ &= a + b\mu_X \end{aligned} \quad \ldots(2.25)$$

Similarly, if $Y = a_1X_1 + a_2 X_2 + a_3 X_3 + \ldots\ldots$

$$E[Y] = E\,[a_1 X_1 + a_2 X_2 + a_3 X_3 + \ldots\ldots]$$

$$\mu_Y = a_1 E[X_1] + a_2E[X_2] + a_3 E[X_3] + \ldots\ldots$$

$$= a_1 \mu_{X_1} + a_2 \mu_{X_2} + a_3 \mu_{X_3} + \ldots\ldots \quad \ldots(2.26)$$

For any random variables X, Y…. Z,

$$E\,(X + Y + \ldots\ldots. + Z) = E(X) + E(Y) + \ldots\ldots.\ E(Z) \quad \ldots(2.27)$$

If the variables are independent

$$E(X \times Y \ldots\ldots\ldots Z) = E(X) \times E(Y) \times \ldots. \times E(Z) \quad \ldots(2.28)$$

Properties of variances:

(i) $\text{Var}\,(a) = 0$ …(2.29)
(ii) $\text{Var}\,(aX) = a^2\,\text{Var}\,(X)$ …(2.30)
(iii) $\text{Var}\,(a + bX) = b^2\,\text{Var}\,(X)$ …(2.31)

2.6.1 Covariance

If X and Y are random variables with their mean μ_X and μ_Y respectively, the covariance (COV) between X and Y is given by

$$\begin{aligned} \text{COV}\,(X, Y) &= E\,[(X - \mu_X)\,(Y - \mu_Y)] \\ &= E\,(XY - \mu_X\,Y - \mu_Y\,X + \mu_X\,\mu_Y) \\ &= E\,(XY) - \mu_X\,E\,(Y) - \mu_Y\,E\,(X) + E\,(\mu_X\,\mu_Y) \end{aligned}$$

or
$$\text{COV}\,(X, Y) = E\,(XY) - E\,(X) \times E\,(Y) \quad \ldots(2.32)$$

However, if X, Y are independent then

$$E\,(XY) = E\,(X) \times E\,(Y)$$

Hence, $\text{COV}\,(X, Y) = 0$

If a function $U = a_1X_1 + a_2X_2 + \ldots\ldots + a_n X_n$ where $X_1, X_2 \ldots. X_n$ are random variables and $a_1, a_2, \ldots. a_n$ are constants then

$$\begin{aligned} E\,(U) &= E\,(a_1 X_1 + a_2X_2 + \ldots\ldots + a_n X_n) \\ &= E\,(a_1X_1) + E(a_2X_2) + \ldots\ldots.. + E(a_n X_n) \\ &= a_1E\,(X_1) + a_2E(X_2) + \ldots\ldots.. + a_n E(X_n) \end{aligned}$$

$$U - E(U) = a_1[X_1 - E\,(X_1)] + a_2[X_2 - E\,(X_2)] + \ldots\ldots.. + a_n[X_n - E\,(X_n)]$$

The Marginal PMF of X is

$$P_X(4) = 0.017 + 0.167 + 0.067 = 0.251$$
$$P_X(5) = 0.05 + 0.267 + 0.133 = 0.450$$
$$P_X(6) = 0.200 + 0.067 + 0.033 = 0.3$$

The marginal PMF of Y is

$$P_Y(2) = 0.017 + 0.05 + 0.200 = 0.267$$
$$P_Y(3) = 0.167 + 0.267 + 0.067 = 0.501$$
$$P_Y(4) = 0.067 + 0.133 + 0.033 = 0.233$$

The conditional PMF of X given Y

$$p_{X|Y}(6|3) = \frac{p_{X,Y}(6,3)}{p_{Y(3)}} = \frac{0.067}{0.501} = 0.134$$

$$p_{X|Y}(5|3) = \frac{p_{X,Y}(5,3)}{p_{Y(3)}} = \frac{0.267}{0.501} = 0.533$$

$$p_{X|Y}(4|3) = \frac{p_{X,Y}(4,3)}{p_{Y(3)}} = \frac{0.167}{0.501} = 0.333$$

The COV(X,Y) and the correlation coefficient $\rho_{X,Y}$ can be estimated by using the equation

$$\rho_{X,Y} = \frac{\text{COV}(X,Y)}{\sigma_X \, \sigma_Y}$$

$$E(X) = [4 \times 0.251 + 5 \times 0.450 + 6 \times 0.3] = 5.054$$

$$\text{Var}(X) = [(4 - 5.054)^2 \times 0.251 + (5 - 5.054)^2 \times 0.450 + (6 - 5.054)^2 \times 0.3] = 0.5486$$

$$\sigma_X = 0.7406$$

Similarly,

$$E(Y) = [2 \times 0.267 + 3 \times 0.501 + 4 \times 0.233] = 2.969$$

$$\text{Var}(X) = [(2 - 2.969)^2 \times 0.267 + (3 - 2.969)^2 \times 0.501 + (4 - 2.969)^2 \times 0.233] = 0.4989$$

$$\sigma_Y = 0.7063$$

$$E(XY) = [4 \times 2 \times 0.017 + 5 \times 2 \times 0.05 + 6 \times 2 \times 0.20 + 4 \times 3 \times 0.167 + 5 \times 3 \times 0.267 + 6 \times 3 \times 0.067 + 4 \times 4 \times 0.067 + 5 \times 4 \times 0.133 + 6 \times 4 \times 0.033] = 14.775$$

$$\text{COV}(X, Y) = E(XY) - E(X)\,E(Y) = [14.775 - (5.054 \times 2.969)] = -0.2303$$

Correlation coefficient, $\rho_{X,Y} = \dfrac{-0.2303}{(0.7407 \times (0.7063)} = -0.44$

Example 2.5 The joint probability mass function of a bi-variate discrete random variable is given by the following Table.

$X \rightarrow$	1	2	3	Total
$Y \downarrow$				
1	0.1	0.1	0.2	0.4
2	0.20	0.3	0.1	0.6
Total	**0.3**	**0.4**	**0.3**	**1.0**

Find (i) the marginal probability mass function of X and Y; (ii) the conditional distribution of X given $Y = 1$; and (iii) $P(X + Y < 4)$.

Solution:

Case (i) From the definition of probability mass function (X, Y),

$$p_X(x_i) = \sum_{y_j} p_{XY}(1, y_j) = \sum_{y_j}^{2} p_{XY}(1, y_j)$$

When $x = 1$, $p_X(x = 1) = \sum_{y_j} p_{XY}(1, y_j)$

$= p_{XY}(1, 1) + p_{XY}(1, 2) = 0.1 + 0.2 = 0.3$

When $x = 2$, $p_X(x = 2) = \sum_{y_j = 1}^{2} p_{XY}(2, y_j) = p_{XY}(2, 1) + p_{XY}(2, 2) = 0.1 + 0.3 = 0.4$

When $x = 3$,

$$p_X(x = 3) = \sum_{y_j = 1}^{2} p_{XY}(3, y_j) = p_{XY}(3, 1) + p_{XY}(3, 2) = 0.2 + 0.1 = 0.3$$

The marginal probability mass function X is

$P_X(x) = 0.3$, when $x = 1$
$= 0.4$, when $x = 2$
$= 0.3$, when $x = 3$.

The marginal probability mass function of Y is given by $p_Y(y_j) = \sum_{x_i} p_{XY}(x_i\, y_i)$

When $y = 1$

$$p_Y(y = 1) = \sum_{x_i}^{3} p_{XY}(x_i, 1) = p_{XY}(1, 1) + p_{XY}(2, 1) + p_{XY}(3, 1) = 0.1 + 0.1 + 0.2 = 0.4$$

When $y = 2$

$$p_Y(y = 2) = \sum_{x_i}^{3} p_{XY}(x_i, 2) = p_{XY}(1, 2) + p_{XY}(2, 2) + p_{XY}(3, 2) = 0.2 + 0.3 + 0.1 = 0.6$$

Example 2.7 The total cost C to manufacture a concrete panel in precast unit is $C = 1.5X + 2Y$ where X is the cost of material, Y is the cost of labour and constants 1.5 and 2 are overhead cost multipliers for material and labour. The cost X and Y are assumed to be uncorrelated random variables with mean values $\mu_X = 5000$/panel and $\mu_Y = 2500$/panel respectively with standard deviation $\sigma_X = 500$/panel, $\sigma_Y = 250$/panel. Find the mean, variance and total cost of manufacture.

Solution: Given cost, $C = 1.5X + 2Y$

$$\begin{aligned}\text{Mean cost, } \mu_C &= E(C) = 1.5E(X) + 2E(Y)\\ &= 1.5\ \mu_X + 2\ \mu_Y\\ &= 1.5(5000) + 2(2500)\\ &= 12500\end{aligned}$$

$$\begin{aligned}\text{Variance of the cost Var}(C) &= \text{Var}\ (1.5X + 2Y)\\ &= (1.5)^2\ \sigma_X^2 + (2)^2\ \sigma_Y^2 + 2(1.5)\ (2)\ \text{COV}(X, Y)\end{aligned}$$

X and Y being uncorrelated random variables, i.e., independent $\text{COV}(X, Y) = 0$

$$\begin{aligned}\text{Var}(C) &= (1.5^2)(500^2) + (2^2)(250^2)\\ \text{Var}(C) &= 812500\end{aligned}$$

$$\begin{aligned}\text{Standard deviation of the cost } \sigma_C &= \sqrt{\text{Var}(C)}\\ &= \sqrt{812500} = 901.39\end{aligned}$$

$$\begin{aligned}\text{The total cost of manufacture } C &= \mu_C \pm \sigma_C\\ &= 12500 \pm 901.39\\ &= 13401.39 \text{ or } 11598.61\end{aligned}$$

2.7 CONTINUOUS VARIABLE

A random variable X can take all possible values between certain limits or in an interval which may be finite or infinite. Some examples of continuous random variables are (i) the strength of concrete; (ii) density of soil; (iii) the length of room; (iv) the wind load; and (v) the angle between two sides. A random variable which can take both discrete and continuous values is called a mixed random variable.

Since a continuous random variable will typically have a zero probability of taking on a specific value, we avoid talking about such probabilities. The CDF can also be used to measure the probability that a random variable takes on a value in certain interval. The probability measures of a continuous random variable is defined by probability density function (PDF), and is given by the following equation defined between intervals a and b

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx \qquad \ldots(2.36)$$

The properties of a cumulative density function (CDF) of variable X if it assumes value between ($-\infty$ and $+\infty$) is given by

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(x)\,dx \qquad \ldots(2.37)$$

If the variable, X assumes value from (0 to ∞), the CDF is given by

$$F_X(x) = P(X \le x) = \int_{0}^{x} f_X(x)\,dx \qquad \ldots(2.38)$$

The derivative of the CDF, $F_X(x)$ gives the PDF of the variable, X.

Hence,
$$f_X(x) = \frac{dF_X(x)}{dx} \qquad \ldots(2.39)$$

The following rules are must for satisfying a CDF, $F_X(x)$

(i) $F_X(-\infty) = 0$ …(2.40)

(ii) $F_X(+\infty) = 1$ …(2.41)

(iii) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ …(2.42)

(iv) $f_X(x) \ge 0$ …(2.43)

(v) $F_X(x) \ge 0$ …(2.44)

The condition $P(a \le X \le b) = \int_{a}^{b} f_X(x)\,dx$ can also be rewritten as

$$P(a \le X \le b) = \int_{-\infty}^{b} f_X(x)\,dx - \int_{-\infty}^{a} f_X(x)\,dx \qquad \ldots(2.45)$$

2.7.1 Expectation of Continuous Random Variable

The continuous random variable X with PDF $f_X(x)$, the expected value of function $g(X)$, of the random variable is given by

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x)\,dx \qquad \ldots(2.46)$$

In general the n^{th} moment of a random variable X is defined as

$$E\left[(X)^n\right] = \int_{-\infty}^{+\infty} (x)^n f_X(x)\,dx \qquad \ldots(2.47)$$

If the random variable is continuous whose probability density function is $f_X(x)$, then

The mean value μ, is $E(X) = \int_{-\infty}^{+\infty} x \times f_X(x)\,dx$ …(2.48)

Variance of X, σ_X^2 is given by $E(X - \mu_X)^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x)\,dx$ …(2.49)

Coefficient of variation, (CoV) = $\frac{\sigma_X}{\mu_X} = \frac{0.4714}{\left(4/3\right)} = 0.35355$

Example 2.9 Find value of K so that the following function can serve as a probability density function.

$$f_X(x) = \begin{cases} 0 & \text{for } x \le 0 \\ Kxe^{-4x^2} & \text{for } x > 0 \end{cases}$$

Solution: For a valid PDF Eq. (2.42) has to be satisfied, i.e.

$\int_{-\infty}^{\infty} f_X(x)dx = 1$, using the boundary conditions we have equation rewritten as

$$\int_0^{\infty} K \cdot x \cdot e^{-4x^2}\, dx = 1$$

Using $u = 4x^2$ the above integral can be expressed as

$$u = 4x^2$$

Now, differentiate w.r.t x

$$\frac{du}{dx} = 8x$$

$$\Rightarrow \qquad du = 8x \times dx$$

$$\therefore \qquad \frac{du}{8x} = dx$$

Now,

$$\int_0^{\infty} K \cdot x \cdot e^{-4x^2} dx = 1$$

$$\Rightarrow \qquad \int_0^{\infty} K \cdot x \cdot e^{-u} \frac{du}{8x} = 1$$

$$\Rightarrow \qquad \frac{K}{8}\int_0^{\infty} e^{-u}\, du = 1$$

$$\text{or} \qquad \frac{K}{8}\left[-e^{-u}\right]_0^{\infty} = 1$$

Hence, $K = 8$ for a valid PDF.

Example 2.10 A random variable X has the following density function.

$$f_X(x) = \begin{cases} Kx^2 & \text{for } -3 \le x \le 3 \\ 0 & \text{elsewhere} \end{cases}$$

Find (a) $P(1 \le x \le 2)$, (b) $P(x \le 2)$, (c) $P(x > 1)$

Solution: The value of K is evaluated first for valid PDF

$\int_{-3}^{3} Kx^2\,dx = 1$. Integrating we get the value of $K = \dfrac{1}{18}$ for a valid PDF

(a) $P(1 \le x \le 2) = \int_{1}^{2} Kx^2\,dx$

$$= \int_{1}^{2} \frac{1}{18}x^2\,dx = \frac{1}{18}\left[\frac{x^3}{3}\right] = \frac{1}{18} \times \frac{1}{3}[2^3 - 1]$$

$$= \frac{7}{54}$$

(b) $P(x \le 2)$ $= \int_{-3}^{2} \frac{1}{18}x^2\,dx$

$$= \frac{35}{54}$$

(c) $P(x > 1)$ $= \int_{1}^{3} \frac{1}{18}x^2\,dx = \frac{18}{18 \times 3} \times [x^3]_1^3$

$$= \frac{26}{54}$$

Example 2.11 Find the CDF for the following PDF of random variable, x.

(a) $f_X(x) = \begin{cases} 6x - 6x^2 & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

(b) $f_X(x) = \begin{cases} \dfrac{x}{4} e^{-\frac{x}{2}} dx & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

Solution: If $f_X(x)$ is a PDF then the CDF is given by $F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx$

(a) $F_X(x) = \int_{-\infty}^{x} (6x - 6x^2)\,dx$

Applying the given boundary conditions

$$= \int_{-\infty}^{0} (6x - 6x^2)\,dx + \int_{0}^{x} (6x - 6x^2)\,dx$$

$$= [0 + (3x^2 - 2x^3)]$$

Hence, CDF is $F_X(x) = (3x^2 - 2x^3)$ for $0 \le x \le 1$

Hence, the PDF can be written as

$$\text{PDF,}\quad f_X(x) = \begin{cases} 0 & \text{for } (x < 1) \\ \dfrac{(x-1)^3}{4} & \text{for } (1 \le x \le 3) \\ 0 & \text{for } (x > 3) \end{cases}$$

Example 2.13 Evaluate the value of K for a proper PDF function given as

$$f_X(x) = Ke^{-\lambda x} \quad \text{for } (x \ge 0)$$

Also find the CDF.

Solution: Using Eq. 2.42, we have

$$\int_0^{\infty} Ke^{-\lambda x}\, dx = 1$$

or $$K = \lambda$$

Hence, PDF is given by $f_X(x) = \lambda e^{-\lambda x}$

The CDF is given by

$$F_X(x) = P(X \le x) = \int_0^x f_X(x)\,dx$$

$$= \int_0^x \lambda e^{-\lambda x}\, dx$$

$$= \left[1 - \frac{1}{e^{-\lambda x}}\right]$$

or $$\text{CDF, } F_X(x) = \left[1 - \frac{1}{e^{-\lambda x}}\right] \quad \text{for } (x \ge 0)$$

Example 2.14 PDF of a random variable is given by $f_X(x) = x^2/9$ for $0 \le x \le 3$, find the mean, median, standard deviation, skewness, and coefficients of variation.

$$\text{Mean } (\mu_X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_0^3 x\frac{x^2}{9}\,dx = \frac{x^4}{36} = 2.25$$

$$\text{Median } (\tilde{\mu}_X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_0^t \frac{x^2}{9}\,dx = 0.5;\ \frac{t^3}{27} = 0.5;\ \text{hence } t = 2.381$$

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f_X(x)\,dx = \int_0^t (x-2.25)^2 \frac{x^2}{9}\,dx = 0.3375$$

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{0.3375} = 0.5809$$

$$\text{Skewness } \mu_3 = \int_{-\infty}^{+\infty} (x-\mu)^3 f_X(x)\,dx = \int_0^t (x-2.25)^3 \frac{x^2}{9}\,dx = -0.168$$

Skewness coefficient $\mu_3 = \frac{\mu^3}{\sigma^3} = -0.857$

Example 2.15 The bearing capacity of soil under a column footing foundation is known to vary between 200 MPa to 400 MPa. The probability density function within this range is given as

$$f_X(x) = \begin{cases} k\left(1 - \frac{x}{400}\right) & \text{for } 200 \text{ MPa} \le x \le 400 \text{ MPa} \\ 0 & \text{elsewhere} \end{cases}$$

If the column is designed to carry a load of 300 MPa. What is the probability of failure of foundation?

Solution:

$$F_X(x) = P(X \le x) = \int_0^x f_X(x)\,dx$$

$$f_X(x) = \int_{200}^{400} k\left(1 - \frac{x}{400}\right) dx = 1.0 = k\left(x - \frac{x^2}{800}\right)\Bigg]_{200}^{400}$$

$$k\left\{\left[400 - \left(\frac{400^2}{800}\right)\right] - \left[200 - \left(\frac{200^2}{800}\right)\right]\right\} = 1.0$$

$$\therefore \qquad k = \frac{1}{50}$$

$$p_s = \int_{200}^{300} k\left(1 - \frac{x}{400}\right) dx = \frac{1}{50}\left[\left(x - \frac{x^2}{800}\right)\right]_{200}^{300} = 0.75$$

The probability of failure of footing is $p_f = 1 - p_s = 1 - 0.75 = 0.25$.

Example 2.16 The probability of operation of a motor continuously without stopping for a period of one year and this random event as success S, $P(S) = 0.98$. The pump stops before the end of one year period as failure F and $P(F) = 0.02$. An engineer using this pump of this kind finds that if pump fails, it will cost ₹ 11.25 lakh owing to the production delays, repair and lost profits $V(F)$ = ₹ 11.25 lakh, if pump does not fail, there are no monetary losses or $V(S)$ = ₹ 0. What is the expected loss or long-term loss. The cost of operating a wastewater treatment system which is a function of its flow $C = 20X^{0.8}$, where C is the annual cost of maintenance in lakhs and the x is the flow in million gallons per day (Mgal/day) and it follows a uniform distribution:

$$f_X(x) = 1/5 \quad \text{for} \quad 5 \le x \le 10 \text{ Mgal/day.}$$

Solution: Expected monetary value (EMV) = $V(S)\,P(S) + V(F)\,P(F)$

$$\text{EMV} = 0 \times 0.98 + 11.25 \times 0.02 = 0.225 \text{ lacs}$$

The conditions that are to be satisfied for a joint distribution are

(i) $p_{X,Y}(x_i, y_j) \geq 0$

(ii) $\sum_{\text{all}\cdot x_i} \cdot \sum_{\text{all}\cdot y_j} p_{X,Y}(x_i, y_j) = 1.00$

(iii) $F_{X,Y}(-\infty, -\infty) = 0$; $F_{X,Y}(\infty, \infty) = 1$

(iv) $F_{X,Y}(-\infty, y) = 0$; $F_{X,Y}(\infty, y) = F_Y(y)$

(v) $F_{X,Y}(x, -\infty) = 0$; $F_{X,Y}(x, \infty) = F_X(x)$

Morever, $F_{X,Y}(x, y)$ is a non-negative, and non-decreasing function of x and y.

If the random variables X and Y are continuous, joint probability density function (PDF) is given by

$$f_{X,Y}(x, y)\, dxdy = P[x \leq X \leq x + dx, y \leq Y \leq y + dy] \quad \text{...(2.59)}$$

Then

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\, dv \cdot du \quad \text{...(2.60)}$$

The joint PDF is also given by

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \quad \text{...(2.61)}$$

The joint PDF is also defined as

$$P(a_1 \leq x \leq a_2, b_1 \leq y \leq b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y) dxdy \quad \text{...(2.62)}$$

Which is the volume under the surface $f(x, y)$

The following conditions are to be satisfied for a proper PDF

(i) $f_{X,Y}(x, y) \geq 0$ for all values of x and y

(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dxdy = 1.00$

(iii) $\int_{a}^{b} \int_{c}^{d} f_{X,Y}(x,y) dxdy = P(a \leq X \leq b, c \leq Y \leq d)$

$$F_{XY}(x, y) = \int_{-\infty}^{x1} \int_{-\infty}^{y1} f_{XY}(x,y) dxdy \quad \text{...(2.63)}$$

(iv) The correlation between two random variables X and Y is defined as

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy\, f_{XY}(x,y) dxdy \quad \text{...(2.64)}$$

If two random variables are uncorrelated with each other, if the correlation between X and Y is equal to the product of their means. $E(XY) = E(X)\,E(Y)$.

The random variables are orthogonal to each other, if the correlation between X and Y is equal to zero.

(v) The covariance between random variables X and Y is then

$$\text{COV}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\mu_X)(y-\mu_Y)\, f_{XY}(x,y)\,dxdy \qquad \ldots(2.65)$$

$$\text{COV}(X, Y) = E(XY) - \mu_X\mu_Y \qquad \ldots(2.66)$$

When COV(X, Y) is zero, the two random variables X and Y are uncorrelated. When the random variables are uncorrelated the correlation is equal to the product of their means.

(vi) The correlation between two random variables X and Y is defined as

$$\rho = \frac{\text{COV}(X,Y)}{\sigma_X\,\sigma_Y} = \frac{E(X,Y) - \mu_X\mu_Y}{\sigma_X\,\sigma_Y} \qquad \ldots(2.67)$$

The correlation coefficient is a number that varies between –1 and +1. When the correlation coefficient is 1 or –1, the variables are perfectly correlated and or have linear relationship between them. When ρ is +1, they are positively correlated or are directly proportional. When ρ is –1 they are negatively correlated or inversely proportional. If correlation is 0, then the random variables are uncorrelated. *If the two random variables are statistically independent then they are uncorrelated and the covariance is zero but converse is not true.*

(vii) The joint probability density function $f_{XY}(x, y)$ of two random variables X and Y, the marginal probability density function of one of the random variables is the integral of the joint probability density function over other random variable.

The marginal probability density function of X is

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dy \qquad \ldots(2.68)$$

The marginal probability density function of Y is

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx \qquad \ldots(2.69)$$

(viii) Conditional probability of two cumulative distribution functions of random variables X and given $Y = y$ is

$$F\left(\frac{X}{Y}\right) = \frac{F_{XY}(x,y)}{F_Y(x)},\ F_Y(y) > 0 \qquad \ldots(2.70)$$

Conditional probability of two cumulative distribution function of random variables Y and given $X = x$ is

The marginal PDF of Y is given by

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x, y)\,dx = \frac{1}{8}\int_0^2 (x+y)\,dx = \frac{(1+y)}{4}$$

(ii) $f_X(x)\, f_Y(y) \neq f_{XY}(x, y)$ hence, X and Y are not indepdent

(iii) Conditional PDF $f_{Y\,|\,X}(y/x)$ is given by

$$f_{Y|X}\left(\frac{y}{x}\right) = \frac{f_{XY}(x, y)}{f_X(x)}; \frac{\frac{1}{8}(x+y)}{\frac{1}{4}(x+1)} = \frac{1\,(x+y)}{2\,(x+1)}\; 0 < x < 2; 0 < y < 2$$

Conditional PDF $f_{X\,|\,Y}\,(x/y)$ is given by

$$f_{X|Y}\left(\frac{y}{x}\right) = \frac{f_{XY}(x,y)}{f_Y(x)}; \frac{\frac{1}{8}(x+y)}{\frac{1}{4}(y+1)} = \frac{1\,(x+y)}{2\,(y+1)}\; 0 < x < 2; 0 < y < 2$$

(iv) $P(0 < Y < 1/2\,|\,X = 1) = \int_0^{1/2} f_{Y|X}(y|x=1)\,dy = \frac{1}{x}\int_0^{1/2} \frac{(1+y)}{2}\,dy = \frac{5}{32}.$

Example 2.18 The joint probability density function of a bi-variate random variable (X, Y)

$$f_{XY}(x, y) = \begin{cases} k(x+y) & \text{for } 0 \leq x1 \\ 0 & \text{for } 0 \leq y \leq 2 \end{cases}$$

Find k and hence ρ_{xy}.

Solution:

$$\int_0^2\int_0^1 k(x+y)\,dxdy = \int_0^2 k\left[\frac{x^2}{2} + xy\right]_0^1 dy = 1 \rightarrow k\int_0^2 \left(\frac{1}{2} + y\right)dy = 1$$

$$k\left[\frac{1}{2}y + \frac{y^2}{2}\right]_0^2 = 1; k = \frac{1}{3}$$

Marginal density function of X is

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)\,dy = \int_0^2 \frac{1}{3}(x+y)\,dy = \frac{1}{3}\left[xy + \frac{y^2}{2}\right]_0^2 = \frac{1}{3}(2x+2)$$

transformed into another random variable Y by the transformation T. The transformation T between X and Y may such that Y is either one-to-one or non-one-to-one transformation.

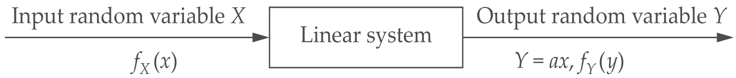

Fig. 2.6 Transformation of $Y = g(X)$

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| \quad \text{or} \quad f_Y(y) = \frac{f_X(x)}{\dfrac{dy}{dx}} \qquad \ldots(2.75)$$

It is more convenient to compute PDF of the new random variable. For the more general case when $Y = g(X)$ has n finite number of real roots, all the transformation which is non-one-to-one, will be broken up into transformations each of which one-to-one and the results are summated together.

Example 2.19 Suppose X is a random variable with quadratic transformation, $Y = 1/X^2$. Find the density function of Y.

Solution: Here the function has two roots for $Y > 0$, $x_1 = -\dfrac{1}{\sqrt{y}}$; $x_2 = \dfrac{1}{\sqrt{y}}$

$$f_Y(y) = \frac{f_X(x_1)}{\dfrac{dy}{dx_1}} + \frac{f_X(x_2)}{\dfrac{dy}{dx_2}}$$

$$\frac{dy}{dx_1} = \frac{-2}{x^3}; \frac{dy}{dx_2} = \frac{-2}{x^3};$$

$$f_Y(y) = \frac{f_X\left(\dfrac{-1}{\sqrt{y}}\right)}{2y^{-3/2}} + \frac{f_X\left(\dfrac{1}{\sqrt{y}}\right)}{2y^{-3/2}}$$

$$f_Y(y) = \frac{y^{3/2}}{2}\left[f_X\left(\frac{-1}{\sqrt{y}}\right) + f_X\left(\frac{1}{\sqrt{y}}\right)\right]$$

2.9.1 Transformation of Two Random Variables

Given two random variables X and Y, suppose we create new random variable U, V, according to some 2×2 transformation. The multivariable calculus states that if a transformation of the form, maps infinitesimal region $A_{x,y}$ to region $A_{u,v}$ then the ratio of the areas of these regions is given by the absolute value of the Jacobian of the transformation.

In other words by Chebyshev's inequality the probability of getting a value which deviates from mean, μ by *at least* $k\sigma$ is *at most* $\left(\frac{1}{k^2}\right)$

or $$P[\,(x-\mu)\geq k\sigma] \leq \left(\frac{1}{k^2}\right) \qquad \ldots(2.79)$$

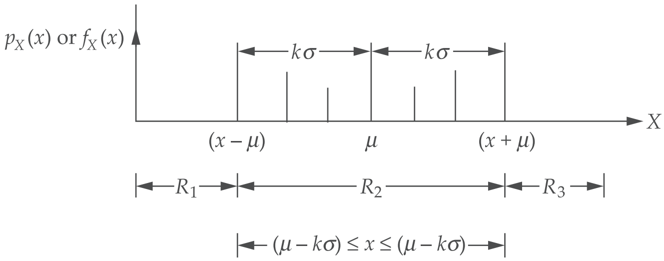

Fig. 2.7 Graphical representation of Chebyshev's inequality

2.10.1 Proof of the Chebyshev's Inequality

Variance of X, is given by

$$\sigma^2 = E(X-\mu)^2$$

$$= \sum_{\text{all}\cdot x_i} (X-\mu)^2 \; p_X(x_i)$$

If the random variable distribution is assumed to be divided in three zones R_1, R_2 and R_3 as in Fig. 2.7, then the expression for variance can be written as

$$\sigma^2 = \sum_{R_1}(X-\mu)^2 \; p_X(x_i) + \sum_{R_2}(X-\mu)^2 \; p_X(x_i) + \sum_{R_3}(X-\mu)^2 \; p_X(x_i) \qquad ...(2.80)$$

where X in region R_1 is given by $X \leq (\mu - k\sigma)$

or $$(X-\mu) \geq k\sigma \qquad ...(2.81)$$

Variable, X in the region R_2 is given by $(\mu - k\sigma) \leq X \leq (\mu + k\sigma)$ ...(2.82)

Similarly, in the region R_3, the variable X is given by the expression

or $$X \geq (\mu + k\sigma) \quad \text{or} \quad (X-\mu) \geq k\sigma \qquad ...(2.83)$$

If the regions R_1 and R_3 are taken then the variance of the distribution shall be greater than the sum of variance of regions R_1 and R_3. Hence, we have the following expression

$$\sigma^2 \geq \sum_{R_1}(X-\mu)^2 \cdot p_X(x_i) + \sum_{R_3}(X-\mu)^2 \cdot p_X(x_i) \qquad ...(2.84)$$

Hence, the probability is at least 0.9375 that strength parameter of the random sample of soil lies between 15 degrees and 35 degrees without knowing the distribution of strength parameter of the soil.

UNSOLVED PROBLEMS

1. Consider the beam *AB* as shown in Fig. 2.8 loaded with two random loads P_1 and P_2. The loads are assumed statistically independent with means, $\mu_{P1} = 4$ kN, and $\mu_{P2} = 6$ kN and standard deviations, $\sigma_{P1} = 0.4$ kN and $\sigma_{P2} = 0.5$ kN. The shear force Q_B and the bending moment M_B at the support *B* are $Q_B = \frac{1}{27}(13P_1 + 23P_2)$, $M_B = \frac{6}{27}(4P_1 + 5P_2)$. Determine the mean, standard deviation, covariance and correlation coefficient between shear force and bending moment.

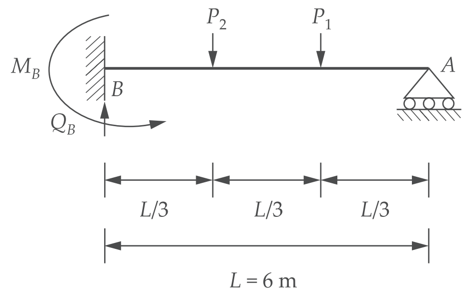

Fig. 2.8

2. A performance function is given by $Y = 3X_1 - 2X_2$ where $\mu_{X_1} = 16.6$; $\mu_{X_2} = 18.8$, $\sigma_{X_1} = 2.45$; $\sigma_{X_2} = 2.83$. The two variables are correlated, and the covariance is equal to 2.0. Obtain the coefficient of variation when X_1 and X_2 are uncorrelated and correlated.
3. Find the CDF for a proper PDF function given as

$$f_X(x) = \begin{cases} 0 & \text{for } x < 1 \\ \frac{1}{2} + \frac{1}{4}(x-3) & \text{for } 1 \le x \le 3 \\ \frac{1}{2} - \frac{1}{4}(x-3) & \text{for } 3 \le x \le 5 \\ 0 & \text{for } x \ge 5 \end{cases}$$

4. A triangular PDF shown in Fig. 2.9 $f_X(x) = \begin{cases} \frac{2}{(b-a)} \frac{(x-a)}{(m-a)} & \text{for } a \le x \le m \\ \frac{2}{(b-a)} \frac{(x-a)}{(m-a)} & \text{for } m \le x \le b \end{cases}$

8. The mean and standard deviation of the force in the members of a truss Fig. 2.11 is given as $E[T_1] = 73.2$ N, $E[T_2] = 51.7$ N and $\sigma_{T1}^2 = 8.784, \sigma_{T2}^2 = 6.204$ and the vertical

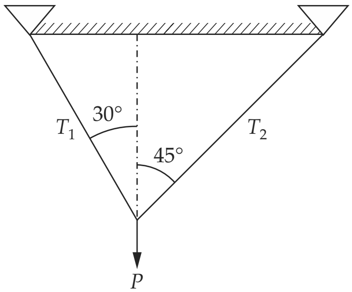

Fig. 2.11

equilibrium $P = T_1 \cos 30° + T_2 \cos 45°$. Estimate the mean value and standard deviation of the load P, given the correlation between T_1, T_2 is zero.