Module: 2 Random Variables

- Introduction -random variables- probability mass Function, distribution and density functions - joint Probability distribution and joint density functions- Marginal, conditional distribution and density functions- Mathematical expectation, and its properties Covariance, moment generating function – characteristic function.

1. ***Random experiment***: It is a process or procedure that produces a set of outcomes, where the outcome is uncertain and cannot be predicted with certainty in advance. Example: Throwing a die, a pack of cards.
2. ***Outcome***: The result of a random experiment will be called an *outcome*.
3. ***Trial***: A trial refers to any single performance of a random experiment. In each trial, one of the possible outcomes occurs. Example: Tossing of a coin: {H, T}
4. **Event:** An event is a specific set of outcomes of a random experiment. An event can consist of one or more outcomes. Events are often what we are interested in when conducting an experiment.
5. ***Exhaustive events or cases***: The complete set of all possible outcomes in a random experiment. In other words, these events cover every possible outcome that could occur in a trial.
6. ***Favourable events or cases***: The number of cases favourable to an event in a trail is the number of outcomes which entail the happening of the event.

Example: In drawing a card from a pack of cards the number of cases favourable to drawing of an ace is 4, for drawing a spade card is 13 and for drawing a red card is 26.

Random Variable

- Intuitively by a *random variable (r.v)* means a real number *X* connected with the outcome of a random experiment *E*.
- Let ***S*** be the sample space associated with a given random experiment **E** .

A real-valued function defined on ***S*** and taking values in $\boldsymbol{R}(-\infty, \infty)$ is called a ***one-dimensional random variable***.

If the function values are ordered pairs of real numbers (*i.e.*, vectors in two-space), the function is said to be a ***two-dimensional random variable***.

More generally, an ***n***-dimensional random variable is simply a function whose domain is ***S*** and whose range is a collection of ***n***-tuples of real numbers (vectors in ***n***-space).

Let us consider the probability space, the triplet $(\boldsymbol{S},\ \boldsymbol{B},\ \boldsymbol{P})$, where $\boldsymbol{S}$ is the sample space, *viz*, space of outcomes, $\boldsymbol{B}$ is the $\boldsymbol{\sigma}$- field of subsets in $\boldsymbol{S}$ and $\boldsymbol{P}$ is a probability function on $\boldsymbol{B}$.

Definition: A random variable is a function $X(\boldsymbol{\omega})$with domain $\boldsymbol{S}$ and range $(-\infty, \infty)$ such that for every real number $\boldsymbol{a}$, the event $[\omega : X(\boldsymbol{\omega}) \leq a] \in \boldsymbol{B}$**.**

Example: Suppose Two (unbiased) coins are tossed $X = number\ of\ heads$.

The sample space S = {HH, HT, TH, TT} $\boldsymbol{P}\{X \leq 1\} = \boldsymbol{P}\{HH, HT, TH\} = \boldsymbol{\frac{3}{4}}$

Note: One-dimensional *r.v.* will be denoted by capital letters, X, Y, Z….etc. A typical outcome of the experiment will be denoted by $\boldsymbol{\omega}$ or *e*.

The values which *X, Y, Z….etc*, can assume are denoted by lower case letters.,

x, y, z…etc.

Discrete Random Variable

If X is a random variable which can take a finite number or countably infinite number of values, then X is called a discrete random variable.

***Example**: Marks* obtained in a test, number of accidents per month, number of telephone calls per unit time, number of successes in *n* trails and so on.

Probability Mass function: If X is a discrete random variable with distinct values $x_1, x_2, \ldots \ldots, x_n$ then the function $\boldsymbol{p_X(x)}$ is defined as :

$$p_X(x_i) = \begin{cases} P(X = x_i) = p_i & \text{if } x = x_i \\ 0 & \text{if } x \neq x_i \; i = 0, 1, 2, \ldots n \end{cases}$$

is called the ***probability mass function*** *or* ***probability function*** of random variable $\boldsymbol{X}$.

The numbers $p(x_i); i = 1, 2, \ldots\ldots$ must satisfy the following conditions:

(i) $p(x_i) \geq 0 \;\; \forall \;\; i$

(ii) $\sum_{i=1}^{n} p(x_i) = 1$

Distribution Function

Definition: Let $\boldsymbol{X}$ be a random variable. The function $\boldsymbol{F}$ defined for all real $\boldsymbol{x}$ by

$$F_X(x) = F(x) = P(X \leq x) = P\{\omega : X(\omega) \leq x\}, -\infty < x < \infty$$

is called the *distribution function* or *cumulative distribution function* of r.v. ($\boldsymbol{X}$).

The domain of the distribution function is $(-\infty, \infty)$ and its range is [0,1]

Properties:

If $\boldsymbol{x}$ is a real number, the set of all $\boldsymbol{\omega}$ in $\boldsymbol{S}$ such that $\mathbf{X}(\boldsymbol{\omega}) = \boldsymbol{x}$ *is*,denoted by $\boldsymbol{X} = \boldsymbol{x}$.

1. $P(X = x) = P(\omega : X(\omega) = x)$

2. $P(X \leq a) = P\{\omega : X(\omega) \in (-\infty, a]\}$

3. $P(a < X \leq b) = P\{\omega : X(\omega) \in (a, b]\} = F(b) - F(a)$

$$P(a \leq X \leq b) = P(X = a) + [F(b) - F(a)]$$

$$P(a < X < b) = F(b) - F(a) - P(X = b)$$

$$P(a \leq X < b) = F(b) - F(a) - P(X = b) + P(X = a)$$

4. $P(X = a \; or \; X = b) = P\{(X = a) \cup (X = b)\}$

5. $P(X = a \; and \; X = b) = P\{(X = a) \cap (X = b)\}$

Properties

If $\boldsymbol{F}$ is the *distribution function* of the *r.v.* $\boldsymbol{X}$ and if $a < b$, then

6. $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$

7. (i) $0 \leq F(x) \leq 1$

(ii) $F(x) \leq F(y)$ if $x < y$

8. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$ and

$$F(\infty) = \lim_{x \to \infty} F(x) = 1$$

Discrete Distribution Function:

A countable number of points $x_1, x_2, x_3, \ldots\ldots x_n$

$p(x_i) \geq 0 \ \forall \ i,$

$\sum_{i=1}^{n} p(x_i) = 1$ such that

$$F(x) = \sum_{i:\, x_i \leq x} p_i = \sum_{x_i} P(X = x_i)$$

If x_i is just integer i, so that $P(X = i) = p_i$; $i = 1, 2, 3, \ldots\ldots$ Then $\boldsymbol{F(x)}$ is a "***step function***" having jump $\boldsymbol{p}$ at i and being constant between each pair of integers.

Properties:

1. $p(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1}),$ where F is the *d.f.* of X.

From a lot of 10 items containing 3 defectives, a sample of 4 items is drawn at random. Let the random variable X denote the number of defective items in the sample. Answer the following when the sample is drawn without replacement.

(i) Find the probability distribution of X.

(ii) Find $P(X \leq 1)$, $P(\mathrm{X} < 1)$ and $P(0 < i < 2)$

Solution: $$P(X = k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$$

where $D =$ number of defective item

$N =$ total number of items

$n =$ number of items drawn in the sample

$k =$ the number of defective items in the sample i)

$$P(X=0)=\frac{\binom{3}{0}\binom{7}{4}}{\binom{10}{4}}=\frac{1\times\binom{7}{4}}{\binom{10}{4}}=\frac{1\times 35}{210}=\frac{35}{210}=\frac{1}{6}$$

$$P(X=1)=\frac{\binom{3}{1}\binom{7}{3}}{\binom{10}{4}}=\frac{3\times\binom{7}{3}}{\binom{10}{4}}=\frac{3\times 35}{210}=\frac{105}{210}=\frac{1}{2}$$

$$P(X=2)=\frac{\binom{3}{2}\binom{7}{2}}{\binom{10}{4}}=\frac{3\times\binom{7}{2}}{\binom{10}{4}}=\frac{3\times 21}{210}=\frac{63}{210}=\frac{3}{10}$$

$$P(X=3)=\frac{\binom{3}{3}\binom{7}{1}}{\binom{10}{4}}=\frac{1\times\binom{7}{1}}{\binom{10}{4}}=\frac{1\times 7}{210}=\frac{7}{210}=\frac{1}{30}$$

ii)

$$P(X\le 1)=P(X=0)+P(X=1)=\frac{1}{6}+\frac{1}{2}=\frac{2}{3}$$

$$P(X<1)=P(X=0)=\frac{1}{6}$$

$$P(0<X<2)=P(X=1)=\frac{1}{2}$$

A random variable X has the following probability function:

Values of $\boldsymbol{X = x}$	0	1	2	3	4	5	6	7
$\boldsymbol{p(X = x)}$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2$+k

(i) Find k, (ii) Evaluate $P(X < 6)$, $P(X \geq 6)$ *and* $P(0 < X < 5)$

(iii) *If* $P\,(X \leq c) > \frac{1}{2}$ find the minimum value of c,

(iv) Determine the distribution function of X.

Solution :

i) Since $\sum_{x=0}^{7} p(x) = 1$

$$k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$$

$$10k^2 + 9k - 1 = 0$$

$$k = -1, \frac{1}{10}$$

$$ii)\ P(X < 6) = P(X = 0) + P(X = 1) + \ldots + P(X = 5)$$

$$= \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} + \frac{1}{100} = \frac{81}{100}$$

$$P(X \geq 6) = 1 - P(X < 6) = \frac{19}{100}$$

$$P(0 < X < 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 8k = \frac{4}{5}$$

iii) $$P(X \leq c) > \frac{1}{2} \Rightarrow c = 4$$

iv)

X	$F_X(x) = P(X \leq x)$
0	0
1	1/10
2	3/10
3	5/10
4	4/5
5	81/100
6	83/100
7	1

Continuous Random Variable

Definition: A random variable X is said to be continuous if it can take all possible values between certain limits.

A continuous random variable is a r.v. that can be measured to any desired degree of accuracy.

Example: Age, height, weight etc.

Probability Density Function (p.d.f)

Definition: Consider the small interval $(\boldsymbol{x}, \boldsymbol{x}+\boldsymbol{dx})$of length $\boldsymbol{dx}$ round the point $\boldsymbol{x}$**.** Let $\boldsymbol{f}(\boldsymbol{x})$ be any continuous function of x so that $\boldsymbol{f}(\boldsymbol{x})\boldsymbol{dx}$ represents the probability that X falls in the infinitesimal interval $(\boldsymbol{x}, \boldsymbol{x}+\boldsymbol{dx})$**.**

$$P(x \leq X \leq x+dx) = f_X(x)dx$$

$$f_X(x) = \lim_{\delta x \to 0} \frac{P(x \ \leq \ X \ \leq \ x+\delta x)}{\delta x}$$

The probability for a variate value to line in the interval dx is $f(x)dx$ and hence the probability for a variate value to fall in the interval $[\alpha, \beta]$ is:

$$P(\alpha \ \leq X \leq \ \beta) = \int_{\alpha}^{\beta} f(x)dx$$

Continuous Distribution Function

Definition: If X is continuous *r.v.* with the *p.d.f.* $f(x)$, then the function

$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt \;\; ; \;\; -\infty < x < \infty$ is called the distribution function (*d.f.*) or sometimes the cumulative distribution function (*c.d.f.*) of the random variable X.

Properties of $\boldsymbol{p.d.f.}$:

(i) $f(x) \geq 0$

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

(iii) The probability $P(E)$given by: $\int_E f(x)dx$ is well defined for any event E.

(iv) $P(X = c) = 0, \ \forall \ c$

$$P(\alpha \leq X \leq \beta) = P(\alpha \leq X < \beta) = P(\alpha < X \leq \beta) = P(\alpha < X < \beta)$$

Properties: 1. $0 \leq F(x) \leq 1 \ ; -\infty < x < \infty$

2. $F'(x) = \frac{d}{dx} F(x) = f(x) \geq 0$

$\Rightarrow$ $F(x)$ is non-decreasing function of x.

3. $F(-\infty) = \lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} \int_{-\infty}^{x} f(x)\, dx = \int_{-\infty}^{-\infty} f(x)\, dx = 0$

$F(+\infty) = \lim_{x \to \infty} F(x) = \lim_{x \to \infty} \int_{-\infty}^{x} f(x)\, dx = \int_{-\infty}^{\infty} f(x)\, dx = 1$

4. $F(x)$ is continuous function of x on the right.

5. The discontinuities of $F(x)$ are at the most countable.

The diameter of an electric cable, say X, is assumed to be a continuous *r.v.* with

$p.d.f : f(x) = 6x(1-x);\ \ 0 \le x \le 1$

(i) Check that $f(x)$ is $p.d.f$,

(ii) Determine a number b such that $P(X < b) = P(X > b)$.

$$\text{i)} \int_0^1 f(x)\,dx = 6\int_0^1 x(1-x)\,dx$$

$$= 6\int_0^1 \left(x - x^2\right) dx = 6\left|\frac{x^2}{2} - \frac{x^3}{3}\right|_0^1 = 1$$

$\therefore f(x)$ is the $p.d.f$ of $r.v. X$

$$P(X < b) = P(X > b)$$

$$\int_0^b f(x)\,dx = \int_b^1 f(x)\,dx$$

$$6\int_0^b x(1-x)\,dx = 6\int_b^1 x(1-x)\,dx$$

$$\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^b = \left[\frac{x^2}{2} - \frac{x^3}{3}\right]_b^1$$

$$4b^3 - 6b^2 + 1 = 0$$

$$\left(b - \frac{1}{2}\right)\left(4b^2 - 4b - 2\right) = 0$$

$$b = \frac{1}{2}, \quad b \simeq 1.366,\ \text{-}0.3660$$

A continuous random variable X has a p.d.f. $f(x) = 3x^2,\ 0 \le x \le 1$. Find a and b such that

i. $P\{X \le a\} = P\{X > a\}$

ii. $P\{X > b\} = 0.05$

i) Since $P(X \le a) = P(X > a)$

$$\Rightarrow P(X \le a) = P(X > a) = \frac{1}{2}$$

$$P(X \le a) = \int_0^a f(x)\,dx = \frac{1}{2}$$

$$a^3 = \frac{1}{2} \Rightarrow a = \left(\frac{1}{2}\right)^{\frac{1}{3}}$$

ii) $P(X > b) = 0.05 \Rightarrow \int_b^1 f(x)\,dx = \frac{1}{20}$

$$1 - b^3 = \frac{1}{20} \Rightarrow b = \left(\frac{19}{20}\right)^{\frac{1}{3}}$$

Let X be a continuous random variable with *p.d.f.* :

$$f(x) = \begin{cases} ax & ; \; 0 \le x \le 1 \\ a & ; \; 1 \le x \le 2 \\ -ax + 3a \, ; & \; 2 \le x \le 3 \\ 0 & ; \; elsewhere \end{cases}$$

(i) Determine the constant $\boldsymbol{a}$.

(ii) Compute $P(X \le 1.5)$.

$$\text{i)} \int_{-\infty}^{\infty} f(x)dx = 1$$

$$\int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{2} f(x)dx + \int_{2}^{3} f(x)dx + \int_{3}^{\infty} f(x)dx = 1$$

$$a = \frac{1}{2}$$

ii) $P\left(X \le 1.5\right) = \int_{-\infty}^{1.5} f\left(x\right)dx = \int_{-\infty}^{0} f\left(x\right)dx + \int_{0}^{1} f\left(x\right)dx + \int_{1}^{1.5} f\left(x\right)dx$

$$= \int_{0}^{1} ax\,dx + \int_{1}^{1.5} a\,dx$$

$$a = \frac{3}{4}$$