Module 3
Random Variables

Introduction to discrete random variables
– Binomial
– Poisson
– Geometric
continuous random Variables
-Normal
-Student’s T
Expectation of random variables, mean and variance.

Random Experiment: A **random experiment** is a process by which we observe something uncertain. After the experiment, the result of the random experiment is known. An **outcome** is a result of a random experiment.

Example: Throwing a die, a pack of cards.

Trial: Any particular performance of a random experiment is called a ***trial***

Sample space: The set of all possible outcomes of a statistical experiment is called the **sample space.** An **event** is a subset of a sample space.

Example: roll a die; sample space: S={1,2,3,4,5,6}. If we would like to know the probability of getting even number, then event is {2,4,6}.

Statistical Definition:

If a random experiment or trail results in ***'n'*** possible outcomes(or cases), out of which ***'m'*** are favourable to the occurrence of an event ***E***, then the probability '**p**' of occurrence of *E*, denoted by ***P*(*E*)**, is given by:

$$p = P\ (E) = \frac{Number\ of\ favourable\ cases}{Total\ number\ of\ exhaustive\ cases} = \frac{m}{n}$$

(i) Since $m \geq 0, n \geq 0 \quad and\ m \leq n$

$P(E) \geq 0 \quad and\ \ P(E) \leq 1 \qquad \Longrightarrow 0 \leq P\ (E) \leq 1$

(ii) $P(E) + P(\bar{E}) = 1$

A real number $\boldsymbol{X}$ connected with the outcome of a random experiment $\boldsymbol{E}$.

For example, if $\boldsymbol{E}$ consists of two tosses the random variable which is the number of **heads** (0,1 or 2).

Outcome	:	HH	HT	TH	TT
Value of $\boldsymbol{X}$	:	2	1	1	0

Thus to each outcome $\boldsymbol{\omega}$, there corresponds a real number $X(\boldsymbol{\omega})$.

Random Variable

Definition: A random variable is a function that associates a real number with each element in the sample space.

Example: Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable: Y, where Y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Note: One-dimensional *r.v.* will be denoted by capital letters, X, Y, Z….etc.

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

Discrete Random Variable

Let $\boldsymbol{X}$ be a finite random variable on a sample space $\boldsymbol{S}$,that is, $\boldsymbol{X}$assigns only finite number or countably infinite number of values to $\boldsymbol{S}$.

$$Say, \qquad R_X = \{x_1, x_2, \dots\dots, x_n, \dots\dots, \infty\}$$

***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.

Then, $\boldsymbol{X}$induces a function $\boldsymbol{p}$ which assigns probabilities to the points in $\boldsymbol{R_X}$ as follows:

$$p(x_i) = p_X(x_i) = P(X = x_i) = P\{s \in S: X(s) = x_i\} \text{ for } i = 1, 2, \dots, n$$

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

p_i is called the ***probability mass 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) $\sum_{i=1}^{n} p(x_i) = 1$

Probability distribution :

The collection of pairs $\{x_i, p_i\}$ is called the probability distribution of the R.V. X.

$X = x_i$	x_1	....	x_i
p_i	p_1	....	p_i

Example:

I have an unfair coin for which $P(H) = p$, where $0 < p < 1$. I toss the coin repeatedly until I observe a heads for the first time. Let Y be the total number of coin tosses. Find the distribution of Y.

Answer:

First, we note that the random variable Y can potentially take any positive integer, so we have $R_Y = \mathbb{N} = \{1, 2, 3, \ldots\}$. To find the distribution of Y, we need to find $P_Y(k) = P(Y = k)$ for $k = 1, 2, 3, \ldots$. We have

$$P_Y(1) = P(Y = 1) = P(H) = p,$$

$$P_Y(2) = P(Y = 2) = P(TH) = (1 - p)p,$$

$$P_Y(3) = P(Y = 3) = P(TTH) = (1 - p)^2 p,$$

$$\begin{array}{cccc} . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{array}$$

$$P_Y(k) = P(Y = k) = P(TT \ldots TH) = (1 - p)^{k-1} p.$$

Thus, we can write the PMF of Y in the following way

$$P_Y(y) = \begin{cases} (1 - p)^{y-1} p & \text{for } y = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

2. If $p = \frac{1}{2}$, find $P(2 \leq Y < 5)$.

Answer:

2. if $p = \frac{1}{2}$, to find $P(2 \leq Y < 5)$, we can write

$$
\begin{aligned}
\mathrm{P}(2 \leq Y < 5) &= \textstyle\sum_{k=2}^{4} P_Y(k) \\
&= \textstyle\sum_{k=2}^{4} (1-p)^{k-1} p \\
&= \tfrac{1}{2}\left(\tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8}\right) \\
&= \tfrac{7}{16}.
\end{aligned}
$$

Continuous Random Variables

A random variable X is said to be a continuous random variable if it takes all possible values between certain limits or in an interval which may be finite or infinite.

Example: operating time between two failures of a computer

Probability density function (p.d.f):

The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$,
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$,
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

Distribution Function (or) Cumulative Distribution Function (cdf)

If X is a R.V, discrete or continuous, then $F(x) = P(X \le x)$ is called the cumulative distribution function of X.

i) If X is discrete then $F(x) = \sum_j p_j$

ii) If X is continuous then $F(x) = P(-\infty \le X \le x) = \int_{-\infty}^{x} f(x)dx$

Relationship between cumulative distribution function and probability distribution function

$$f(x) = \frac{dF(x)}{dx}$$

If $p(x)$ is the p.m.f of a random variable 'X' then

$$Mean = \sum_{n=0}^{\infty} xp(x)$$

$$\text{Variance} = \sum_{n=0}^{\infty} (x - mean)^2 p(x)$$

$$\text{Moment about the mean} = \sum_{n=0}^{\infty} (x - mean)^r p(x)$$

If $f(x)$ is the p.d.f of a random variable 'X' which defined in the interval (a, b) then

(i) $\text{Mean} = \int_a^b xf(x)dx$ (ii) $\text{Variance} = \int_a^b (x - mean)^2 f(x)dx$

(iii) $\text{Moment about mean} = \int_a^b (x - mean)^r f(x)dx$

Taxonomy of Probability Distributions

Discrete probability distributions

- Binomial distribution
- Multinomial distribution
- Poisson distribution
- Hypergeometric distribution

Continuous probability distributions

- Normal distribution
- Standard normal distribution
- Gamma distribution
- Exponential distribution
- Chi square distribution
- Lognormal distribution
- Weibull distribution

Discrete Probability Distributions

Binomial Distribution

- In many situations, an outcome has only two outcomes: success and failure.
- An experiment when consists of repeated trials, each with such outcomes is called Bernoulli process. Each trial in it is called a Bernoulli trial.

Example: Firing bullets to hit a target.

- Suppose, in a Bernoulli process, we define a random variable $X \equiv$ the number of successes in trials.
- Such a random variable obeys the binomial probability distribution, if the experiment satisfies the following conditions:

1)The experiment consists of n trials.

2)Each trial results in one of two mutually exclusive outcomes, one labelled a *"success"* and the other a *"failure"*.

3)The probability of a success on a single trial is equal to $\boldsymbol{p}$. The value of p remains constant throughout the experiment.

4)The trials are independent.

Defining Binomial Distribution

Definition: **Binomial distribution**

The function for computing the probability for the binomial probability distribution is given by

$$f(x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

for x = 0, 1, 2, …., n

Here, $f(x) = P(X = x)$, where X denotes "the number of success" and $X = x$ denotes the number of success in x trials.

If X follows binomial distribution with parameters n and p symbolically, we express $X \sim B(n, p)$ or $b(n, p)$.

Moment Generating Function

$$
\begin{aligned}
M_x(t) &= E[e^{tx}] \\
&= \sum_{x=0}^{n} e^{tx} p(x) \\
&= \sum_{x=0}^{n} e^{tx} n_{c_x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} n_{c_x} (pe^t)^x q^{n-x}
\end{aligned}
$$

$$
\left[As\ (x+a)^n = n_{c_0} x^n + n_{c_1} x^{n-1} a + n_{c_2} x^{n-2} a^2 + \cdots \ldots \ldots + n_{c_n} a^n\right]
$$

$$
= (Pe^t + q)^n
$$

Mean and Variance of Binomial distribution

$$
\begin{aligned}
\mu_1' &= \left[\frac{d}{dt}\{M_x(t)\}\right]_{t=0} \\
&= \left[\frac{d}{dt}\{(pe^t + q)^n\}\right]_{t=0} \\
&= [n(pe^t + q)^{n-1}pe^t]_{t=0} \\
&= n(p + q)^{n-1}p \qquad\qquad [\because p + q = 1] \\
&= np
\end{aligned}
$$

Mean of the Binomial distribution is np.

$$\mu_2' = \left[\frac{d^2}{dt^2}\{M_x(t)\}\right]_{t=0}$$

$$= \left[\frac{d^2}{dt^2}\{(pe^t + q)^n\}\right]_{t=0}$$

$$= \left[\frac{d}{dt}\{n(pe^t + q)^{n-1}pe^t\}\right]_{t=0}$$

$$= np[e^t(n-1)(pe^t + q)^{n-2}pe^t + (pe^t + q)^{n-1}e^t]_{t=0}$$

$$= np[(n-1)(p+q\ )^{n-2}p + (p+q\ )^{n-1}] \qquad [\because p + q = 1]$$

$$= np[(n-1)p + 1]$$

$$\mu_2' = n(n-1)p^2 + np$$

$$
\begin{aligned}
\sigma^2 &= \mu_2' - (\mu_1')^2 \\
&= n(n-1)p^2 + np - n^2p^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np(1-p) \\
&= npq
\end{aligned}
$$

Variance of Binomial distribution is npq

Ex: The mean of a binomial distribution is 5 and standard deviation is 2. Determine the distribution.

Answer:

$$Given\ mean\ = \mathrm{n}p = 5$$

$$\mathrm{S.D} = \sqrt{npq} = 2$$

$$\mathrm{npq} = 4$$

$$5\mathrm{q} = 4$$

$$\mathrm{q} = 4/5$$

$$\mathrm{p} = 1 - \mathrm{q} = 1/5$$

$$\mathrm{n}(1/5) = 5$$

$$\mathrm{n} = 25$$

Hence the Binomial distribution is

$$p(X = x) = n_{C_x} p^x q^{n-x} = 25_{C_x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{25-x}, x = 0,1,2 \ldots \ldots 25$$

Ex: The mean and variance of a binomial distribution are 4 and 4 / 3. Find $P(X \geq 1)$.

Answer:

Given mean $np = 4$, Variance $npq = 4/3$

$$\frac{npq}{np} = \frac{4/3}{4} = \frac{1}{3}$$

$$\mathbf{q} = 1/3; \mathrm{p} = 2/3$$

$$\mathrm{np} = 4$$

$$\mathrm{n}(2/3) = 4$$

$$\mathrm{n} = 6$$

$$p(X \geq 1) = 1 - P(X < 1)$$

$$= 1 - P(X = 0)$$

$$= 1 - 6_{C_0} p^0 q^{6-0}$$

$$= 1 - 6_{C_0} (2/3)^0 (1/3)^{6-0}$$

$$= 1 - (1/3)^6 = 0.998$$

Ex: An irregular 6 faced die is such that the probability that it gives 3 even numbers in 5 throws is twice the probability that it gives 2 even numbers in 5 throws. How many sets of exactly 5 trials can be expected to give no even number out of 2500 sets?

Answer:

Let X denote the number of even numbers obtained in 5 trials.
Given $P(X = 3) = 2\ P(X = 2)$.

$$5_{C_3} p^3 q^2 = 2 * 5_{C_2} p^2 q^3$$
$$p = 2q = 2(1 - p) = 2 - 2p$$
$$3p = 2$$

Hence $p = 2/3$, and $q = 1/3$.

Now, P(getting no even number)=$P(X = 0)$=$5_{C_0} p^0 q^5 = (1/3)^5 = \frac{1}{243}$

Number of sets having no success (no even number) out of N sets =$N * \ P(X = 0)$
Required number of sets =2500* 1 / 243=10, nearly.

Problem 1:A die is thrown 4 times. Getting a number greater than 2is a success.
Find the probability of getting (i) exactly one success, (ii)less than 3 successes.

Problem 2:If the chance that any one of 5 telephone lines is busy at any instant is 0.01, what is the probability that all the lines are busy? What is the probability that more than 3 lines are busy?

Problem 3:A die is thrown three times. Getting a ”3” or a ”6” is considered to be success. Find the probability of getting at least two successes.

Problem 4:If 20% of the bolts produced by a machine are defective, determine the probability that out of 4 bolts chosen at random (i) 1,(ii) 0 will be defective.

Problem 5:Out of 1000 families of 3 children each, how many families would you expect to have two boys and one girl, assuming that boys and girls are equally likely.

Problem 6:The average percentage of failures in a certain examination is 40. What is the probability that out of a group of 6 candidates, at least 4 pass in the examination?

Problem 7: X follows a binomial distribution such that $4P(X=4) = P(X=2)$. If $n=6$, find p the probability of success.

Problem 8:Find the maximum n such that the probability of getting no head in tossing a coin n times is greater than 0.1.

Problem 9:If the sum of the mean and variance of a binomial distribution of 5 trials is 95, find the binomial distribution.