Probability Distributions

Random Variables(RV)

- A variable whose values depends on the outcomes of a random phenomenon is called a random variable.
- Act whose outcome cannot be predicted in advance is called an random experiment.
- Outcome of a random experiment is called an event.
- Set of possible values that X can take is called the support of X (χ).
- Sample space is the set of all outcomes of the experiment.
- Sum of probability of all possible values of RV is 1 always.

- Types of Random Variable :

There are two types of random variable :

1) Discrete Random Variable
2) Continuous Random Variable

Discrete Random Variable : A random variable X is said to be discrete, if the total number of values X can take is finite, i.e. the support of X is either finite or countable.

Probability mass function : It is a function p of the discrete random variable. It is a function p:X->R as follows :

$$P(a) = \begin{cases} p_i & a=a_i \quad i=0,1,2,3,.... \\ 0 & \text{otherwise} \end{cases}$$

Probability distribution : It is the list of values of X_i and their corresponding probabilities P_i.

For example : If we want the probability of getting number of heads in three tosses of a coin.

Let RV, X be "getting number of heads".

Total number of outcomes=8

S={HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}

In this case x=0,1,2,3

P(getting no heads)=P(X=0)=1/8

P(getting 1 heads)=P(X=1)=3/8

P(getting 2 heads)=P(X=2)=3/8

P(getting 3 heads)=p(X=3)=1/8

Probability distribution :

X_i	0	1	2	3
P_i	1/8	3/8	3/8	1/8

Cumulative distribution function : Another function which plays an important role in random variable is cumulative distribution function. Cumulative distributive function (c.d.f) F: R->[0,1] of the random variable Xis defined as

$F(b) = P(X \leq b)$, for $-\infty < b < \infty$.

important properties of the c.d.f. F (.) are :

(a) F (b) is a non-decreasing function of b.

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

(c) $\lim_{b \to -\infty} f(b)=0$

1) A box contains twice as many blue marbles as redmarbles. One marble is drawn at random from the box and is replaced ; then a second marble is drawn at random from the box. If both marbles are red you win Rs. 50 ; if both marbles are blue you loose Rs. 10 ; and if they are of different colour then neither you loose nor you win. Determine the probability distribution for the amount you win or loose?

No. of red marbles = x

No. of blue marbles=2x

Total no. of marbles= 3x

P(getting red marble)=x/3x=1/3

P(getting blue marble)=2x/3x=2/3

Let X be the random variable "the amount we win or loose".

If both marbles are red you win Rs. 50

X=+50

P(X=50)=(Red,Red)=1/3*1/3=1/9

If both marbles are blue you loose Rs. 10

X=-10

P(X=-10)=(Blue,Blue)=2/3*2/3=4/9

If they are of different colour then neither you loose nor you win.

X=0

P(X=0)=P((Red,Blue)or (Blue,Red))

=2/3*1/3+1/3*2/3

=2/9+2/9

=4/9

Probability distribution

X	0	50	-10
P(X)	4/9	1/9	4/9

Binomial Distribution

Binomial distribution is also known as Bernoulli's distribution. It is used with the experiments where there are only two possible outcomes.

Characteristics of a binomial distribution :

- Fixed number of trials.
- Each trial is independent of the others.
- Each trial has two outcomes.
- Probability of each outcome remains constant from trial to trial.

Formula :

$P(X=x) = {}^{n}C_{x}\, p^{X} q^{n-x}$ where n is total number of outcomes, p is probability of success and q is probability of failure.

${}^{n}C_{x}=n!/(n-r)!r!$

Few applications of binomial distribution :

- A newborn is a girl or a boy
- Tossing a coin : it has only 2 outcomes head or tail
- Writing a true/False test
- Winning a lottery: it has only 2 outcomes win or loose
- Germination of seeds : it has only 2 outcomes will germinate or not.

1. An unbiased dice is thrown 3 times. If getting 2or 5 is considered a success. Find the probability of atleast 2 success.

Ans : Let X be a binomial RV of “getting 2 or 5”.

$n=3$

P(getting 2 or 5)=1/6+1/6

=2/6

=1/3 =P(success)=p

$q=1-1/3$

$=2/3$

Probability of atleast 2 success is :

$P(X\geq 2)=P(X=2)+P(X=3)$

$P(X\geq2)=P(X=2)+P(X=3)$

$P(x)={}^{n}C_{x}p^{x}q^{n-x}$

$P(X=2)={}^{3}C_{2}(1/3)^{2}(2/3)^{1}$

$=(3!/2!)(1/3)2(2/3)$

$=3*1/9*2/3$

$=12/81$

$P(X=3)={}^{3}C_{3}(1/3)^{3}(2/3)^{0}$

$=1*1/27*1$

$=1/27$

$P(X\geq2)=12/81+1/27=15/81$

2. If we want the probability of getting number of heads in three tosses of a coin.

i. No heads

ii. 1 heads

iii. 2 heads

iv. 3 heads

Let RV, X be “getting number of heads”.

n=3,p=1/2,q=1/2

i. No heads

$P(X=0)={}^{3}C_{0}(1/2)^{0}(1/2)^{3}$

$=1*1*1/8$

$=1/8$

ii. 1 head

$P(X=1)={}^{3}C_{1}(1/2)^{1}(1/2)^{2}$

$=3*1/2*1/4$

$=3/8$

iii. 2 heads

$P(X=2)= {}^{3}C_{2}(1/2)^{2}(1/2)^{1}$

$=3*1/4*1/2$

$=3/8$

iv. 3 heads

$P(X=3)= {}^{3}C_{3}(1/2)^{3}(1/2)^{0}$

$=1*1/8*1$

$=1/8$

Poisson Distribution

It is the limiting case of binomial distribution. It is used to find the probability of an experiment in a given time interval or specified region of space.

Characteristics of Poisson Distribution :

a) The average occurrence rate per unit time is constant.

b) Occurrence in an interval is independent of what has happened previously.

c) The chance that more than one occurrence will happen at the same time is negligible.

Formula :

$P(x)=(e^{-m} * m^{x})/x!$

m=np, where n: total number of outcomes , p: probability of success

Few examples of Poisson distribution :

- Number of typing errors on a page
- Traffic flow
- Number of accidents on a particular stretch of read in a week
- Number of telephone calls to a call center

1. Calls at a particular call center occur at an average rate of 8 calls per 10 minutes. Suppose that the operator leaves his position for a 5 minute coffee break. What is the chance that exactly one call comes in while the operator is away?

m=8*5/10=4

$P(X=m)=e^{-m} * m^x/x!$

$=e^{-4} * 4^1/1!$

=0.073

Continuous Random Variable

A continuous random variable is a random variable where the data can take infinitely many values. A random variable measuring the time taken for something to be done is continuous since there are an infinite number of possible times that can be taken.

Check your Progress 1

1. Suppose you take a 50-question multiple-choice examination, guessing your answer, and are interested in the number of correct answers obtained. Then (a) What is the random variable X that you will consider for this situation? (b) What is the set of possible values of X in this example? (c) What does P(X=10) mean in this context?

a) Let RV X be the number of correct answers obtained.[Because we have to find the number of correct answers.]

b) X=0,1,2,3,.....50

c) P(X=10) means the probability that the number of correct answers is 10.

Check your Progress 2

1. Which of the variables given below are discrete? Give reasons for your answer.

(a)The daily measurement of snowfall at Shimla. (b) The number of industrial accidents in each month in West Bengal. (c) The number of defective goods in a shipment of goods from a manufacturer.

(a) The daily measurement of snowfall at Shimla. This is because this is for an interval, so it is a continuous variable.

(b) The number of industrial accidents in each month in West Bengal. The number of accidents in an industry is finite, so it is a discrete variable.

(c) The number of defective goods in a shipment of goods from a manufacturer. The number of defective goods in a shipment is finite, so it is a discrete variable.

Check your Progress 3

1. A farmer buys a quantity of cabbage seeds from a company that claims that approximately 90% of the seeds will germinate if planted properly. If four seeds are planted, what is the probability that exactly two will germinate?[**very important**]

Let X be a binomial RV "seeds that will germinate".

$n=4$

$p=90/100=0.9$

$q=1-0.9$

$=0.1$

P(exactly two will germinate)=P(X=2)

$P(X=2)={}^4C_2(0.9)^2(0.1)^2$

$=6*0.81*0.01$

$=0.0486$

Check your Progress 4

1. If a bank receives on an average $\lambda = 6$ bad Cheques per day, what is the probability that it will receive 4 bad checks on any given day.[very important]

Let X be a Poisson RV, "bank receives bad cheque".

$\lambda = 6$

$P(X=4)=e^{-6} * 6^4/4!$

$=(0.0024*1296)/24$

$=(0.0025*1296)/24$

$=0.135$