STATISTICS

Introduction

Random Experiment: Even though all the possible outcomes of an experiment are known before the experiment, the outcome of a particular performance of the experiment cannot be predicted owing to the number of unknown causes, such an experiment is called a random experiment.
Example: Tossing of a coin.

Random Variable: It is a function that assigns a real number X(s) to every element $s \in S$ corresponds to the random experiment E.

Example:
Rolling a die: 6 faces of a die X=1 or X=2, X= 3,X=4 or 5 or 6.

Tossing 3 coins {HHH, HTT,THH, …}

Event: Outcome of a random experiment is called an event. A random variable can be either discrete or continuous. A random variable is said to be discrete if the set of values defined by it over the sample space is finite or countably infinite or non-negative integer where as continuous random variable can assume any real value in the interval say (a,b)

If the random variable X is a discrete random variable, then the probability function P(X) is called the probability mass function and if the random variable is a continuous one then P(X) is called probability density function.

Probability Distributions: It is function which gives the probabilities of occurrence of different possible outcomes of a random experiment. There are two types of distribution functions, namely discrete distribution and continuous distribution. Binomial and Poisson are discrete distributions and Normal distribution is a continuous distribution.

Binomial distribution:

Let us consider a trial in which there two possible outcomes, one is success denoted by *p* and other is failure denoted by *q* (equals *1- p*). As the probability of success and failure are mutually exclusive, its sum is 1. If this trial is repeated "*n*" number of times independently and in this '*n*' trails if the *x* trials denotes the successes , then *(n-x)* trials denotes the failure. That is if the probability of success is p then that p is repeating x times and *q(=1-p)* appears *(n-x)* times So for each way of getting this x successes is given by $p^x.q^{(n-x)}$. So for getting these successes in *n*- trials are $n_{c_x} p^x q^{n-x}$. And it is called probabixlity density function of the variate x.

i.e., $P(X=x)=n_{c_x} p^x q^{n-x}$ where $0 \le p, q \le 1$ and $x = 0, 1, 2, 3 \ldots. n$. The density function also called Probability mass function as *x* takes the discrete values.

Mean and **variance of Binomial distribution:**

$$Mean = E(x) = \sum_{x=0}^{n} x\, P(X=x) = \sum_{x=0}^{n} x \left({}^{n}c_{x}\, p^{x} q^{n-x} \right)$$

$$= np \sum_{x=1}^{n} \left({}^{n-1}c_{x-1}\, p^{x-1} (1-p)^{n-1-(x-1)} \right)$$

$$= np \left(\overline{1-p} + p \right)^{n-1} = np = \mu$$

$$Variance = \sigma^2 = \sum_{x=0}^{n} (x-\mu)^2\, P(X=x) = \sum_{x=0}^{n} (x-\mu)^2\, {}^{n}c_{x}\, p^{x} q^{n-x}$$

$$= \sum_{x=0}^{n} x^2\, {}^{n}c_{x}\, p^{x} q^{n-x} - n^2 p^2$$

Consider

$$\sum_{x=0}^{n} x^2\, {}^{n}c_{x}\, p^{x} q^{n-x} = \sum_{x=1}^{n} \left(x(x-1) + x \right) \frac{n!}{(n-x)!\; x!} p^{x} q^{n-x}$$

$$= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(n-x)!(x-2)!} p^{x-2} q^{n-x} = n(n-1)p^2 (q+p)^{n-2} + np$$

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

$$\therefore \sigma^2 = n(n-1)p^2 + np - n^2 p^2 = np(1-p) = npq$$

Mode is the value of x for which P(X=x) is maximum.

When $p = 1/2$, the mode is $\frac{n}{2}$ if n is even and the modes are $\frac{n-1}{2}$ and $\frac{n+1}{2}$ if n is odd.

1. **Out of 10000 families with 4 children each, find the number of families all whose children are daughters.**

   If p denotes the probability of a female child, then the probability of a male child is q. Also the probabilities of female and male child are equally likely.

   $\therefore$ The Probabilities of female and male are p = ½ and q = ½.

   n = 4 and Number of family = 10000 = N
   If x denotes the probability of a female child,
   Then the number of families whose children are daughters = $N \times P(x=4)$

   $$P(x=4) = 4_{c_4}\left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

   $$\therefore \textit{The number of families whose children are daughters} = 10000 \times \frac{1}{16} = 625$$

2. If the sum of the mean and variance of a Binomial distribution for 5 trials is 1.8. Find the binomial distribution.

   The mean of the binomial distribution = μ = np

   The variance of the binomial distribution = σ^2 = npq

   Here   n = 5. Also q= 1 – p .

   Given   Mean + variance = 1.8

   i.e., n p + n p q = 1.8

n p + n p (1-p) = 1.8 $\quad\quad\quad\quad$ ($\because$ q = 1- p)

$$\Rightarrow 5\times p+5\times p-5\times p^2 = 1.8$$

The above equation is quadratic in p, on solving the quadratic equation, p value can be determined. Then the value is $p=\dfrac{10\pm\sqrt{100-36}}{10}$ or p = 1.8, 0.2.

As probability value cannot exceed 1, p = 0.2.

$\therefore$ Mean $=\mu=$ 5 x 0.2 = 1 and the variance $=\sigma^2=$ 5 x 0.2 x 0.8 = 8

The distribution function is $5_{c_x}(0.2)^x(0.8)^{n-x}$

Exercise

1. The Probability that a pen manufactured by a company will be defective is1/10. If 12 such pens are manufactured find the probability that

   (a) Exactly 2 will be defective (b) at least two will be defective (iii) none will be defective.

   Ans: 0.2301, 0.3412 0.2833

2. In 256 sets of 12 tosses of a coin, in how many cases one can expect 8 heads and 4 tails ?

   Ans: 30.9 or approximately 31.