**Probability Distributions**

**Scope**

* Meaning of probability distributions
* Binomial Distribution
* Poisson Distribution

**Theory**

Probability distribution as the name suggests is the listing of all possible outcomes of an experiment together with their respective probabilities.

For example, if we toss a fair coin two times, the following is the list of all possible outcome of this experiment and their respective probabilities

|  |  |
| --- | --- |
| Outcome | Probability |
| TT | ¼ |
| TH | ¼ |
| HT | ¼ |
| HH | ¼ |

Stated in other way, the probability distribution of the number of heads obtained in these two tosses of the coin is given as follows:

|  |  |
| --- | --- |
| Number of heads (X) | Probability P(X) |
| 0 | ¼ |
| 1 | ½ |
| 2 | ¼ |
|  | 1.0 |

Probabilities distributions describe the probabilities of the events of a random variable. The events can be measured either as discrete values or as continuous values.

For example, if a quality control inspector examines four radios randomly from a production lot, and the number of defective radios from this sample can be presented by the variable X, then is X is a random variable with the following possible discrete values

{0, 1, 2, 3, 4}

So, Probability distributions are classified either as discrete or continuous depending upon the nature of the variable being considered.

Examples of discrete distributions are *Binomial Distribution* and *Poisson Distribution.*

On the other hand, the term continuous distribution implies that it can only be measured to some pre-determined degree of accuracy.

Time, weight or distances are all measured on a continuous scale. The value of a continuous variable cannot be precise at any one point and hence it takes any one value in an interval. The most frequently used continuous probability distribution is the N*ormal Distribution.*

**Binomial Distribution**

**Theory**

It is one of the simplest and most frequently used discrete probability distribution and is very useful in many practical situations involving *either/or* types of events.

It describes the distribution of probabilities where there are only two mutually exclusive outcomes for each trial of an experiment. For example, when checking the quality of a product, we can see that either the item is *good* or it is *defective*.

Events are described as either *success* or *failure.*  Success is simply the outcome in which we are interested. If we are interested in the probability of a head in the toss of a coin, then the outcome tail will be considered a failure.

The probability of success is symbolised by *p* and the probability of failure is symbolised by *q,* which is also (1-p).

*Some conditions*

Each trial is independent of other trials. The outcome of one trial does not affect the outcome of any other trial.

The probability of success *p* remains constant from trial to trial. So is the probability of failure.

**Illustration**

Suppose we toss a coin five times. The outcome head is designated as success and outcome tail as failure, with probability of success and failure being respective (p) and (q). Suppose that the following was the sequence of outcomes of these tosses.

H, T, H, T, T

This means that there are two successes and three failures in the given order as above. The join probability of the mutually exclusive events is given as:

p x q x p x q x q = p2q3

However we are not concerned with any particular sequence of the outcome. We want two successes in any order out of the give tosses. Then it can be shown that there are ten different ways of obtaining two heads out of five tosses. By applying addition rule of probability, we can see that the probability of getting any sequence with two heads and three tails would be ten times the probability of the single sequence obtained above.

How did we get 10?

Apply combination formula.

nCx = n!/x! (n-x)!

Hence, the general formula of calculating the probability of x successes out of n trials is given by

P(x) = (n x) pxq(n-x)

P(2) = (5 2) (0.5)2(0.5)3

0.3125

**Applications in the real world- Example**

If a new drug is found to be effective 40% of the time, then what is the probability that in a random sample of 4 patients, it will effective on 2 of them?

Solution

Let’s define effective as success and non-effective as failure.

Then:

p = 0.4 and q = 0.6

x = 2, n = 4

Now,

P(x) = (4 2)(0.4)2(0.6)(4-2)

0.3456

**Mean and Standard Deviation of a Binomial Distribution**

Mean (mu) = np

SD (sigma) = sqrt(npq)

**Example**

In a manufacturing process, a packaging machine produces 5% defective packages. Find the mean and the standard deviation of the number of defective packages in a random sample of 60 packages.

**Solution**

Mean = np

Variance = root(npq)

In our case,

P= 0.05, n = 60, q = (1-p) = 0.95

Hence,

Mean = 60 x 0.05 = 3

S.D = root(npq) = root(60 x 0.05 x 0.95) = 1.69

**Poisson distribution**

**Theory**

Poisson distribution is another theoretical discrete distribution which is useful for modelling certain real situations. It differs from the binomial distribution in the sense that in the binomial distribution we must be able to count the number of successes and number of failures, while in Poisson distribution, all we want to know is the average number of successes in a given unit of time or space.

In many situations, it is not possible to count the number of failures even though we can know the number of successes.

For example, in the case of patients coming to the hospital for emergency treatment, we can always count the number of patients arriving in any given hour. If the number of patients arriving is considered a as the number of successes, then we cannot know the number of failures because it is not possible to count the number of patients *not* coming for emergency treatment in that hour.

Accordingly, it is not possible to determine the total number of possible outcomes (successes and failures) and hence binomial distribution cannot be applied as a decision making tool. In such a situation, we can use Poisson distribution, if we know the average number of patients arriving for emergency treatment per hour. It is assumed that such arrival of patients is a random phenomenon and hence the exact number of patients arriving in any hour is not predictable.

**Applications**

* Telephone calls going through a switchboard system
* Number of cars passing through the India Gate
* Number of customers coming to a bank for services

**Formula**

We need to know the average number of events per unit of time. The symbol for this average is Lambda.

It could be average number of cars passing under a bridge in any given hour or it could be the average number of machine breakdowns per month or it could be the average number of customers arriving at a bank per day and so on.

The probability that exactly x events will occur in a given time is given as

P(x) = lambdax e-lambda/x!

**Example**

Assume that on an average 3 persons enter the bank for service every 10 minutes. What is the probability that exactly 5 customers will enter the bank in a given 10 minute period, assuming that the process can be described by Poisson distribution?

Solution

P(x) = 35 (2.17828)-3/5!

=0.1008

**Example**

Customers arrive at a photocopying machine at an average rate of two every 10 minutes. The number of arrivals is distributed according to a Poisson distribution. What is the probability that:

1. There will be no arrivals during any period of ten minutes
2. There will be exactly one arrival during this time period
3. There will be more than two arrivals during this time period

Solution

P(0) = .1353

P(1) = .2707

P(2) = .2707

More than two arrivals is

P(x>2) = 1- [P(0) + P(1) + P(2)]

=1 – [.1353+.2707+.2707]

=.3233