**Notes – Ch 5 Discrete Probability Distribution**

**Random Variable -** A numerical description of the outcome of an experiment. A variable is random if it takes on different values as a result of the outcomes of a random experiment.

**Discrete Random Variable -** A random variable that may assume either a finite number of values or an infinite sequence of values.

**Continuous Random Variable -** A random variable that may assume any numerical value in an interval or collection of intervals

**Probability Distribution -** A rule that assigns probabilities to different values of the random variable

**Discrete Probability Distribution -** A function, denoted by f(x), that provides the probability that x assumes a particular value for a discrete random variable.

In the development of probability function for any discrete random variable, the following two conditions must be satisfied

f(x) ≥ 0

∑f(x) = 1

These conditions must hold because the f(X = x) values are probabilities. First condition specifies that all probabilities must be greater than or equal to zero. For the second condition, we note that for each value x, f(x) = f(X = x) is the probability of the event that the random variable equals x. Since by definition, all x means all the values the random variable X may take, and since X may take on only one value at a time, the occurrences of these values are mutually exclusive events, and one of them must take place. Therefore, the sum of all the probabilities f(X = x) must be 1

**Expected Value -** A measure of the central location of a random variable. The expected value is a weighted average of the values the random variable where the weights are the probabilities. The expected value does not have to be a value the random variable can assume.

**Variance -** The variance is a weighted average of the squared deviations of a random variable from its mean. The weights are the probabilities.

**Standard Deviation -** The positive square root of the variance.

The standard deviation is measured in the same units as the random variable and therefore is often preferred in describing the variability of a random variable. The variance σ2 is measured in squared units and is thus more difficult to interpret.

**Discrete uniform probability distribution -** A probability distribution for which each possible value of the random variable has the same probability

f(x) = 1/n where n = the number of values the random variable may assume

**Binomial Probability distribution**

**Binomial experiment** - The binomial distribution can be used to determine the probability of x successes in n trials whenever the experiment has the following properties:

1. The experiment consists of a sequence of n identical trials.
2. Two outcomes are possible on each trial, one called success and the other failure. Sample space of all possible outcomes is {Success, Failure}
3. The probability of a success p does not change from trial to trial. Consequently, the probability of failure, 1 – p or q, does not change from trial to trial.
4. The trials are independent. i.e. The outcome of any trial or sequence of trials do not affect the outcomes of subsequent trials.

**Binomial probability distribution:** A probability distribution showing the probability of x successes in n trials of a binomial experiment.

**Binomial probability function** - The function used to compute binomial probabilities.

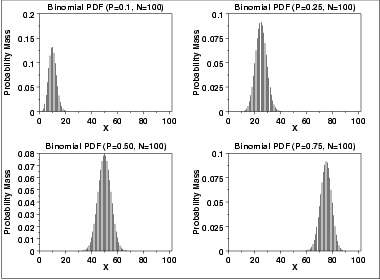
f(x) = for x = 1, 2, 3, ……..n

**Mean:** E(x) = μ = np

**Variance:** Var(x) = σ2 = np(1-p)

**Uses:**

* A political poll of voters is carried out. Each polled voter is asked whether or not they currently approve of the Prime Minister.
* A random sample of schools is obtained. The schools are assessed on their compliance with a suitable policy on sun exposure for their students.
* A random sample of police personnel are interviewed. Each person is assessed as to whether or not they show appropriate awareness of different cultures.
* A random sample of drivers are drug tested, and it is recorded whether or not they are positive for recent methamphetamine use.
* A random sample of footballers is chosen and their record of injuries assessed, according to whether or not they have had more than 3 episodes of concussion.

**Observations:**

* The shape and location of binomial distribution changes as p changes for a given p. As p increase for a fixed n, the binomial distribution shifts to the right.
* If p and q are unequal, the distribution is skew. If p is less than ½ the distribution is positively skewed and when p is more than ½ the distribution is negatively skewed
* If n is large and if neither p nor q is too close to zero, the binomial distribution can be closely approximated to normal distribution.

**Cases where Bernoulli process fails:** If the machine has wear and tear, the probability of producing defective parts keeps on changing. The Bernoulli condition for binomial fail. When humans are involved such as in interviews each trial may not be independent.

**Poisson Probability distribution -** The Poisson distribution is used when it is desirable to determine the probability of obtaining *x* occurrences over an interval of time or space.

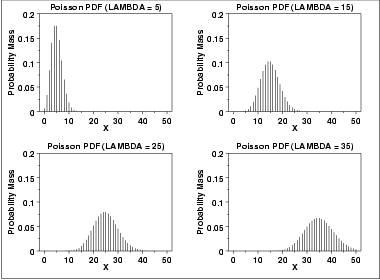
The following assumptions are necessary for the Poisson distribution to be applicable:

1. The probability of an occurrence of the event is the same for any two intervals of equal length.
2. The occurrence or non-occurrence of the event in any interval is independent of the occurrence or non-occurrence of the event in any other interval.

**Poisson probability distribution** - A probability distribution showing the probability of x occurrences of an event over a specified interval of time or space.

**Poisson probability function -** The function used to compute Poisson probabilities.

**Properties:**

1. It is skewed to right. This is the reason why the Poisson distribution has been called the distribution of rare events (The probabilities tend to be high for small number of occurrences****
2. Mean = Variance

**Uses -** Poisson distribution may be expected in situations where the chance of occurrence of any event is small, and we are interested in the occurrence of the event and not in its non-occurrence. Poisson Distribution is applicable in situations where events occur at random points of time and space wherein our interest lies only in the number of occurrences of the event.

For example, number of road accidents, number of defective items, number of deaths in flood or because of snakebite or because of a rare disease etc. In these situations, we know about the occurrence of an event although its probability is very small, but we do not know how many times it does not occur. For instance, we can say that two road accidents took place today, but it is almost impossible to say as to how many times, accident fails to take place. The reason is that the number of trials is very large here and the nature of event is of rare type. The Poisson random variable X, counts the number of times a rare event occurs during a fixed interval of time or space.

The patients for a service at a health center, the arrival of trucks and cars at tollbooth, number of printing mistakes in a book

**Poisson as an approximation of Binomial Distribution –** Poisson can be used to approximate Binomial if

**Hypergeometric Probability distribution:**

Conditions which characterize the distribution:

1. The result of each draw can be classified into one of the two categories
2. The probability of a success changes on each draw
3. Successive draws are dependent
4. The drawing is repeated a fixed number of times

**Hypergeometric probability distribution:** A population of size N has r successes and N-r failures. From this population we select a sample of size n. Hypergeometric probability distribution shows the probability of x successes in n trials from a population of size N with r successes and N - r failures.

**Hypergeometric probability function:** The function used to compute hypergeometric probabilities.

Like the binomial, it is used to compute the probability of x successes in n trials. But, in contrast to the binomial, the probability of success changes from trial to trial. For binomial the sample is drawn with replacement from a finite population or without replacement from an infinite population. For hypergeometric the sample data are drawn without replacement for a finite population. When n increases without limit, the distribution approaches binomial. Hence binomial probabilities may be used as approximation to hypergeometric probabilities where σ/n is small.

The mean and variance of hypergeometric distribution is as follows:

Mean = E(x) =

Variance =