Chapter 2

RELIABILITY AND SIX SIGMA – PROBABILITIC MODELS

Is man one of God's blunders or is God one of man's blunders?
- Friedrich Nietzsche

2.1 INTRODUCTION

Development of any product involves designing processes that can result in the desired product. Successful development of a product depends on the effectiveness of the underlying processes. Naturally, process variability is a concern for any product development team, since it affects the reliability and quality of the product. To manage the processes, we have to measure it, especially the variations and the uncertainties associated with them. Apart from various other factors, market success of a product would depend on how the company manages to meet the customer requirements under uncertainties. For example, customers may be happy with a courier company which delivers documents with mean duration of two days and standard deviation of one day compared to another company which delivers the documents with mean duration of one day with standard deviation of two days. In the first case, the probability that a document will not be delivered within 4 days is approximately 2.27% (assuming normal distribution), whereas the same probability is 6.67% in case of second company, although its average time to deliver the document is one day. Similarly, a customer purchasing a product such as consumer durables would expect the product to function properly to begin with and maintain the functionality for a minimum duration. This 'minimum duration' is subjective and would vary from product to product. Here, there are two uncertainties; one corresponds to the probability that a newly purchased product will function, and two the product will maintain the functionality for a minimum period of time. The

former corresponds to Six Sigma and the latter is related to the reliability of the product.

Probability theory is the fundamental building block for both reliability theory and Six Sigma. In this chapter we introduce the basic concepts in probability theory which are essential to understand the rest of the book. This chapter is not intended for a rigorous treatment of all-relevant theorems and proofs in probability theory, but to provide an understanding of the main concepts in probability theory that can be applied to problems in reliability and Six Sigma.

2.2 PROBABILITY TERMINOLOGY

In this section we introduce various terminologies used in probability that are essential for understanding the rudiments of probability theory. To facilitate the discussion some relevant terms and their definitions are introduced.

Experiment

An experiment is a well-defined act or process that leads to a single well-defined outcome. Figure 2.1 illustrates the concept of random experiments. Every experiment must:

1. Be capable of being described, so that the observer knows when it occurs.
2. Have one and only one outcome.

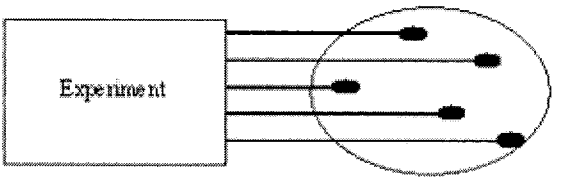

Figure 2-1. Graphical Representation of an Experiment and its outcomes

Elementary event

An elementary event is every separate outcome of an experiment, also known as a sample point.

From the definition of an experiment, it is possible to conclude that the total number of elementary events is equal to the total number of possible outcomes.

Sample space

The set of all possible distinct outcomes for an experiment is called the sample space for that experiment.

Usually, the symbol S is used to represent the *sample space*, and small letters, *a, b, c,..,* for elementary events that are possible outcomes of the experiment under consideration. The set S may contain either a finite or infinite number of elementary events. Figure 2.2 is a graphical presentation of the sample space.

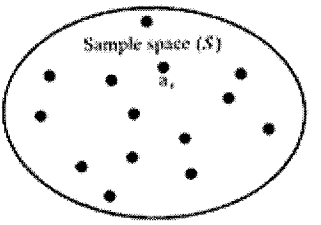

Figure 2-2. Graphical Presentation of the Sample Space

Event

Event is a subset of the sample space, that is, a collection of elementary events.

Capital letters A, B, C,... are usually used for denoting events. For example, if the experiment performed is measuring the speed of passing cars at a specific road junction, then the elementary event is the speed measured, whereas the sample space consists of all the different speeds one might possibly record. All speed events could be classified in to, say, four different speed groups: *A* (less than 30 km/h), *B* (between 30 and 50 km/h), *C* (between 50 and 70 km/h) and *D* (above 70 km/h). If the measured speed of the passing car is, say 35 km/h, then the event *B* is said to have occurred.

2.3 ELEMENTARY THEORY OF PROBABILITY

The theory of probability is developed from axioms proposed by the Russian mathematician *Kolmogrov*. All other rules and relations in probability theory are derived from axioms of probability.

2.3.1 Axioms of probability

In cases where the outcome of an experiment is uncertain, it is necessary to assign some measure that will indicate the chances of occurrence of a particular event. Such a measure of events is called the *probability of the event* and symbolized by *P(.)*, (*P(A)* denotes the probability of occurrence of event *A*). The function which associates each event *A* in the sample space *S*, with the probability measure *P(A)*, is called the *probability function* - the probability of that event. A graphical representation of the probability function is given in Figure 2.3.

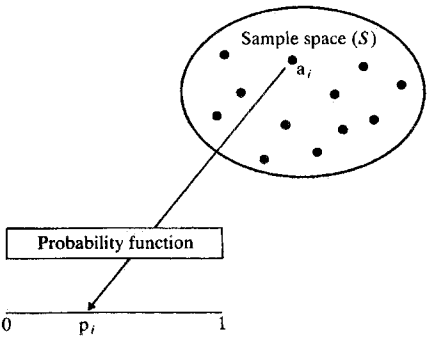

Figure 2-3. Graphical representation of probability function

Formally, the probability function is defined as:

A function which associates with each event A, a real number, P(A), the probability of event A, such that the following axioms are true:

1. P(A) > 0 for every event A,
2. P(S) = 1, (probability of the sample space is equal to one)
3. The probability of the union of mutually exclusive events is the sum of their probabilities, that is

$$P(A_1 \cup A_2 ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$$

In essence, this definition states that each event A is paired with a non-negative number, probability P(A), and that the probability of the sure event S, or P(S), is always 1. Furthermore, if A_1 and A_2 are any two mutually exclusive events (that is, the occurrence of one event implies the non-occurrence of the other) in the sample space, then the probability of their union $P(A_1 \cup A_2)$, is simply the sum of their two probabilities, $P(A_1) + P(A_2)$.

2.3.2 Rules of probability

The following elementary rules of probability are directly deduced from the original three axioms, using set theory:

a) For any event A, the probability of the complementary event, denoted by A', is given by

$$P(A') = 1 - P(A) \tag{2.1}$$

b) The probability of any event must lie between zero and one:

$$0 \leq P(A) \leq 1 \tag{2.2}$$

c) The probability of an empty or impossible event, ϕ, is zero.

$$P(\phi) = 0 \tag{2.3}$$

d) If occurrence of an event A implies occurrence of an event B, so that the event class A is a subset of event class B, then the probability of A is less than or equal to the probability of B:

$$P(A) \leq P(B) \tag{2.4}$$

e) In order to find the probability that A or B or both occur, the probability of A, the probability of B, and also the probability that both occur must be known, thus:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2.5}$$

f) If A and B are mutually exclusive events, so that $P(A \cap B) = 0$, then

$$P(A \cup B) = P(A) + P(B) \tag{2.6}$$

g) If n events form a partition of the sample space S, then their probabilities must add up to one:

$$P(A_1) + P(A_2) + \ldots + P(A_n) = \sum_{i=1}^{n} P(A_i) = 1 \tag{2.7}$$

2.3.3 Joint event

Any event that is an intersection of two or more events is called a joint event.

There is nothing to restrict any given elementary event from the sample space from qualifying for two or more events, provided that those events are not mutually exclusive. Thus, given the event A and the event B, the joint event is $A \cap B$. Since a member of $A \cap B$ must be a member of set A, and also of set B, both A and B events occur when $A \cap B$ occurs. Provided that the elements of set S are all equally likely to occur, the probability of the joint event could be found in the following way:

$$P(A \cap B) = \frac{\text{number of elementary events in A} \cap \text{B}}{\text{total number of elementary events}}$$

2.3.4 Conditional probability

If A and B are events in a sample space which consists of a finite number of elementary events, the conditional probability of the event B given that the event A has already occurred, denoted by $P(B|A)$, is defined as:

$$P(B|A) = \frac{\text{P(A} \cap \text{B)}}{\text{P(A)}}, \qquad P(A) > 0 \tag{2.8}$$

The conditional probability symbol, $P(B|A)$, is read as the probability of occurrence of event B given that the event A has occurred. It is necessary to satisfy the condition that $P(A)>0$, because it does not make sense to consider

the probability of B given A if event A is impossible. For any two events A and B, there are two conditional probabilities that may be calculated:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad \textit{and} \qquad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

(The probability of B, given A) (The probability of A, given B)

One of the important applications of conditional probability is due to Bayes theorem, which can be stated as follows:

If $(A_1, A_2, \ldots, A_N)$ represents the partition of the sample space (N mutually exclusive events), and if B is subset of $(A_1 \cup A_2 \cup \ldots \cup A_N)$, as illustrated in Figure 2.4, then

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1)+\ldots+P(B|A_i)P(A_i)+\ldots+P(B|A_N)P(A_N)} \qquad (2.9)$$

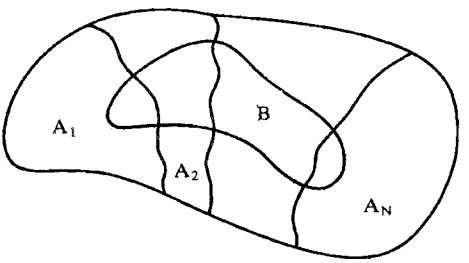

Figure 2-4. Graphical Presentation of the Bayes Theorem

2.4 PROBABILITY DISTRIBUTION

Consider the set of events $A_1, A_2, \ldots, A_n$, and suppose that they form a partition of the sample space S. That is, they are mutually exclusive and exhaustive. Then the corresponding set of probabilities, $P(A_1), P(A_2), \ldots, P(A_n)$, is a probability distribution. An illustrative presentation of the concept of probability distribution is shown in Figure 2.5.

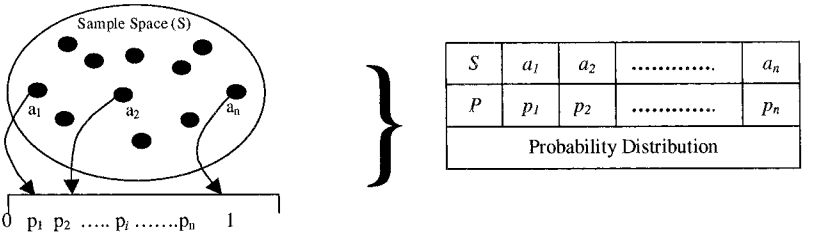

Figure 2-5. Graphical representation of Probability Distribution

2.5 RANDOM VARIABLE

A function that assigns a number (usually a real number) to each sample point in the sample space S is a random variable.

Outcomes of experiments may be expressed either in numerical or non-numerical terms. However, In order to compare and analyze them it is convenient to use real numbers. So, for practical applications, it is necessary to assign a numerical value to each possible elementary event in a sample space *S*. Even if the elementary events themselves are already expressed in terms of numbers, it is possible to reassign a unique real number to each elementary event. The function that achieves this is known as *the random variable*. In other words, a random variable is a real-valued function defined on a sample space. Usually it is denoted with capital letters, such as *X*, *Y* and *Z*, whereas small letters, such as *x, y, z, a, b, c*, and so on, are used to denote particular values of a random variables, see Figure 2.6.

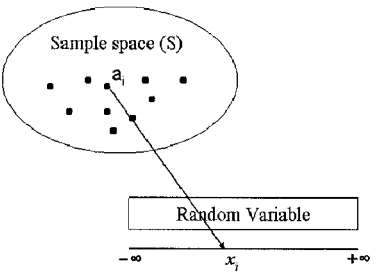

Figure 2-6. Graphical Representation of Random Variable

If X is a random variable and r is a fixed real number, it is possible to define the event A to be the subset of S consisting of all sample points 'a' to which the random variable X assigns the number r, $A = (a : X(a) = r)$. On the other hand, the event A has a probability $p = P(A)$. The symbol p can be interpreted, generally, as the probability that the random variable X takes value r, $p = P(X = r)$. Thus, the symbol $P(X = r)$ represents the probability function of a random variable. Therefore, by using the random variable it is possible to assign probabilities to real numbers, although the original probabilities were only defined for events of the set S, as shown in Figure 2.7.

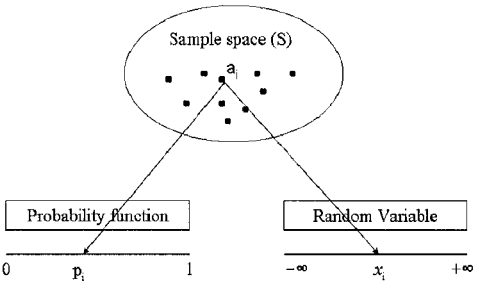

Figure 2-7. Relationship between probability function and a random variable

The probability that the random variable X, takes value less than or equal to certain value 'x', is called the *cumulative distribution function, F(t).* That is,

$$P[X \leq x] = F(x)$$

2.5.1 Types of random variables

Depending on the values assumed, random variables can be classified as discrete or continuous. The main characteristics, similarities and differences for both types are briefly described below.

Discrete random variables

If the random variable X can assume only a particular finite or countably infinite set of values, it is said to be a discrete random variable.

There are very many situations where the random variable X can assume only a *finite* or *countably infinite* set of values; that is, the possible values of X are finite in number or they are infinite in number but can be put in a one-to-one correspondence with a set of real number.

Continuous random variables

If the random variable X can assume any value from a finite or an infinite set of values, it is said to be a continuous random variable.

Let us consider an experiment, which consists of recording the temperature of a cooling liquid of an engine in the area of the thermostat at a given time. Suppose that we can measure the temperature exactly, which means that our measuring device allows us to record the temperature to any number of decimal points. If X is the temperature reading, it is not possible for us to specify a finite or countably infinite set of values. For example, if one of the finite set of values is 75.965, we can determine values 75.9651, 75.9652, and so on, which are also possible values of X. What is being demonstrated here is that the possible values of X consist of the set of real numbers, a set which contains an infinite (and uncountable) number of values. Continuous random variables have enormous utility in reliability since the random variables time to failure is a continuous random variable.

2.6 THE PROBABILITY DISTRIBUTION OF A RANDOM VARIABLE

Taking into account the concept of the probability distribution and the concept of a random variable, it could be said that the probability distribution of a random variable is a set of pairs:

$$\{r_i, P(X = r_i),\ i = 1,n\} \tag{2.10}$$

Figure 2.8 shows the relationship between a random variable and its probability distribution. The easiest way to present this set is to make a list of all its members. If the number of possible values is small, it is easy to specify a probability distribution. On the other hand, if there are a large number of possible values, a listing may become very difficult. In the extreme case where we have an infinite number of possible values (for example, all real numbers between zero and one), it is clearly impossible to make a listing. Fortunately, there are other methods that could be used for specifying a probability distribution of a random variable:

1. Functional method, where a specific mathematical function exists from which the probability of random variable taking any value or interval of values can be calculated.
2. Parametric method, where the entire distribution is represented through one or more parameters known as summary measures.

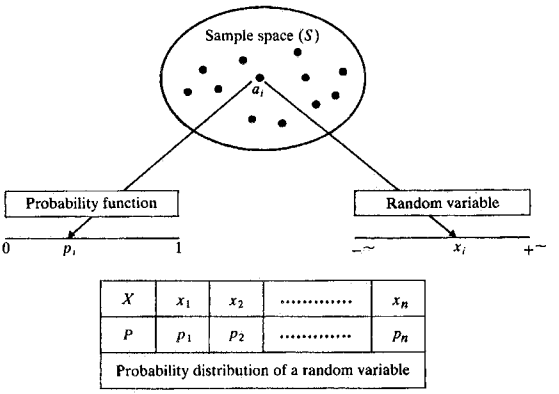

Figure 2-8. Probability Distribution of a Random Variable

2.6.1 Functional method

By definition, a function is a relation where each member of the domain is paired with one member of the range. In this particular case, the relation between numerical values that random variables assume and their probabilities will be considered. The most frequently used functions for the description of probability distribution of a random variable are the

probability mass function, the probability density function, and the cumulative distribution function. Each of these will be analyzed and defined in the remainder of this chapter.

Probability mass function

This function is related to a discrete random variable and it represents the probability that the discrete random variable, X, will take one specific value x_i, $p_i = P(X = x_i)$. Thus, a probability mass function, which is usually denoted as $PMF(.)$, places a mass of probability p_i at the point x_i on the X-axis.

Given that a discrete random variable takes on only n different values, say $a_1, a_2, \ldots, a_n$, the corresponding $PMF(.)$ must satisfy the following two conditions:

$$\begin{aligned} &1. \quad P(X = a_i) \geq 0 \qquad\qquad \textit{for } i = 1,2,\ldots,n \\ &2. \quad \sum_{i=1}^{n} P(X = a_i) = 1 \end{aligned} \tag{2.11}$$

In practice this means that the probability of each value that X can take must be non-negative and the sum of the probabilities must be 1. Thus, a probability distribution can be represented by a pair of values (a_i, p_i), where $i = 1,2,\ldots,n$, as shown in Figure 2.9. The advantage of such a graph over a listing is the ease of comprehension and a better provision of a notion for the nature of the probability distribution.

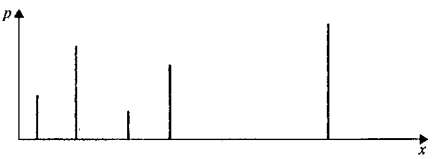

Figure 2-9. Probability Mass Function

Probability density function

In the previous section discrete random variables were discussed in terms of probabilities $P(X = x)$, the probability that the random variables take an *exact* value. However, consider the example of an infinite set. For a specific type of car, the volume of the fuel in the fuel tank is measured with

only some degree of accuracy. What is the probability that a car selected at random will have *exactly* 16 litres of fuel? This could be considered as an event that is defined by the interval of values between, say 15.5 and 16.5, or 15.75 and 16.25, or any other interval $16 \pm 0.1i$, where i is very small, but not exactly zero. Since the smaller the interval, the smaller the probability, the probability of exactly 16 litres is, in effect, zero.

In general, for continuous random variables, the occurrence of any exact *value of X may be regarded as having zero probability.*

The Probability Density Function, $f(x)$, which represents the probability that the random variable will take values within the interval $x \le X \le x + \Delta(x)$, where $\Delta(x)$ approaches zero, is defined as:

$$f(x) = \lim_{\Delta(x) \to 0} \frac{P(x \le X \le x + \Delta(x))}{\Delta x} \tag{2.12}$$

As a consequence, the probabilities of a continuous random variable can be discussed only for *intervals* that the random variable X can take. Thus, instead of the probability that X takes on a specific value, say *'a'*, we deal with the so-called *probability density* of X at *'a'*, symbolized by $f(a)$. In general, the probability distribution of a continuous random variable can be represented by its *Probability Density Function, PDF*, which is defined in the following way:

$$P(a \le X \le b) = \int_a^b f(x)dx \tag{2.13}$$

A fully defined probability density function must satisfy the following two requirements:

$$f(x) \ge 0 \qquad \text{for all } x$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

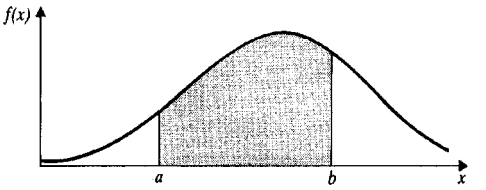

Figure 2-10. Probability Density Function for a Hypothetical Distribution

The *PDF* is always represented as a smooth curve drawn above the horizontal axis, which represents the possible values of the random variable X. A curve for a hypothetical distribution is shown in Figure 2.10 where the two points a and b on the horizontal axis represent limits which define an interval. The shaded portion between *'a'* and *'b'* represents the probability that X takes a value between the limits *'a'* and *'b'*.

Cumulative distribution function

The probability that a random variable X takes a value at or below a given number 'a' is often written as:

$$F(a) = P(X \leq a) \tag{2.14}$$

The symbol $F(a)$ denotes the particular probability for the interval $X \leq a$. This function is called the *Cumulative Distribution Function, CDF*, and it must satisfy certain mathematical properties, the most important of which are:

1. $0 \leq F(x) \leq 1$
2. *if* $a < b$, $F(a) \leq F(b)$
   $F(\infty) = 1$ *and* $F(-\infty) = 0$

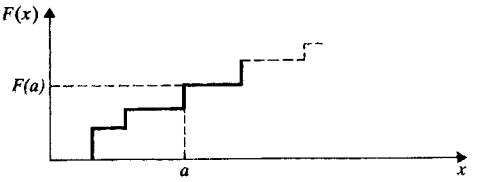

Figure 2-11. Cumulative Distribution Function for Discrete Variable

In general, the symbol $F(x)$ can be used to represent the cumulative probability that X is less than or equal to x. For the discrete random variable, it is defined as:

$$F(a) = \sum_{i=1}^{n} P(X = x_i) \tag{2.15}$$

Whereas in the case of continuous random variables it will take the following form:

$$F(a) = \int_{-\infty}^{a} f(x)dx \tag{2.16}$$

Hypothetical cumulative distribution functions for both types of random variable are given in Figures 2.11 and 2.12.

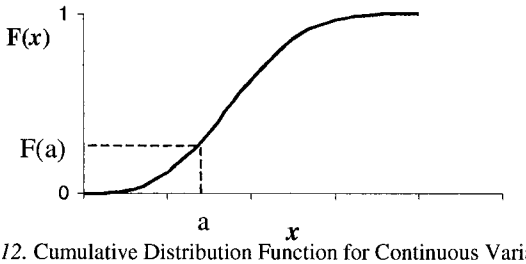

Figure 2-12. Cumulative Distribution Function for Continuous Variable

2.6.2 Parametric method

In some situations it is easier and more efficient to look only at certain characteristics of distributions rather than to attempt to specify the distribution as a whole. Such characteristics summarize and numerically describe certain features for the entire distribution. Two general groups of such characteristics applicable to any type of distribution are:

a) *Measures of central tendency* (or location) which indicate the typical or the average value of the random variable.
b) *Measures of dispersion* (or variability) which show the spread of the difference among the possible values of the random variable.

In many cases, it is possible to adequately describe a probability distribution with a few measures of this kind. It should be remembered, however, that these measures serve only to summarize some important features of the probability distribution. In general, they do not completely describe the entire distribution.

One of the most common and useful summary measures of a probability distribution is *the expectation* of a random variable, *E(X)*. It is a unique value that indicates a location for the distribution as a whole (In physical science, expected value actually represents the centre of gravity). The concept of expectation plays an important role not only as a useful measure, but also as a central concept within the theory of probability and statistics.

If a random variable, say X, is discrete, then its expectation is defined as:

$$E(X) = \sum_{x} x \times P(X = x) \tag{2.17}$$

Where the summation is over all values the variable X can take. If the random variable is continuous, then the expectation is defined as:

$$E(X) = \int_{-\infty}^{+\infty} x \times f(x)dx \tag{2.18}$$

For a continuous random variable the expectation is also defined as:

$$E(X) = \int_{-\infty}^{+\infty} [1 - F(x)]dx \tag{2.19}$$

If c is a constant, then

$$E(cX) = c \times E(X) \tag{2.20}$$

Also, for any two random variables X and Y,

$$E(X + Y) = E(X) + E(Y)$$

2.6.2.1 Measures of central tendency

The most frequently used measures are mean, median and mode.

The mean of a random variable is simply the expectation of the random variable under consideration. Thus, for the random variable, X, the mean value is defined as:

$$Mean = E(X) \tag{2.21}$$

The median, is defined as the value of X which is midway (in terms of probability) between the smallest possible value and the largest possible value. The median is the point, which divides the total area under the *PDF* into two equal parts. In other words, the probability that X is less than the median is $1/2$, and the probability that X is greater than the median is also $1/2$. Thus, if $P(X \leq a) = 0.50$ and $P(X \geq a) = 0.50$ then *'a'* is the *median* of the distribution of X. In the continuous case, this can be expressed as:

$$\int_{-\infty}^{a} f(x)dx = \int_{a}^{+\infty} f(x)dx = 0.50 \tag{2.22}$$

The mode, is defined as the value of X at which the *PDF* of X reaches its highest point. If a graph of the *PMF (PDF)*, or a listing of possible values of X along with their probabilities is available, determination of the mode is quite simple.

A central tendency parameter, whether it is mode, median, mean, or any other measure, summarizes only a certain aspect of a distribution. It is easy to find two distributions which have the same mean but are not similar in any other respect.

2.6.2.2 Measures of dispersion

The mean is a good indication of the location of a random variable, but it is possible that *no single value of the random variable may match with mean*. A deviation from the mean, D, expresses the measure of error made by using the mean as measure of the random variable:

$$D = x - M$$

Where, x, is a possible value of the random variable, X. The deviation can be taken from other measures of central tendency as well, such as the median or mode. It is quite obvious that the larger such deviations are from

a measure of central tendency, the more the individual values differ from each other, and the more apparent the spread within the distribution becomes. Consequently, it is necessary to find a measure that will reflect the spread, or variability, of individual values.

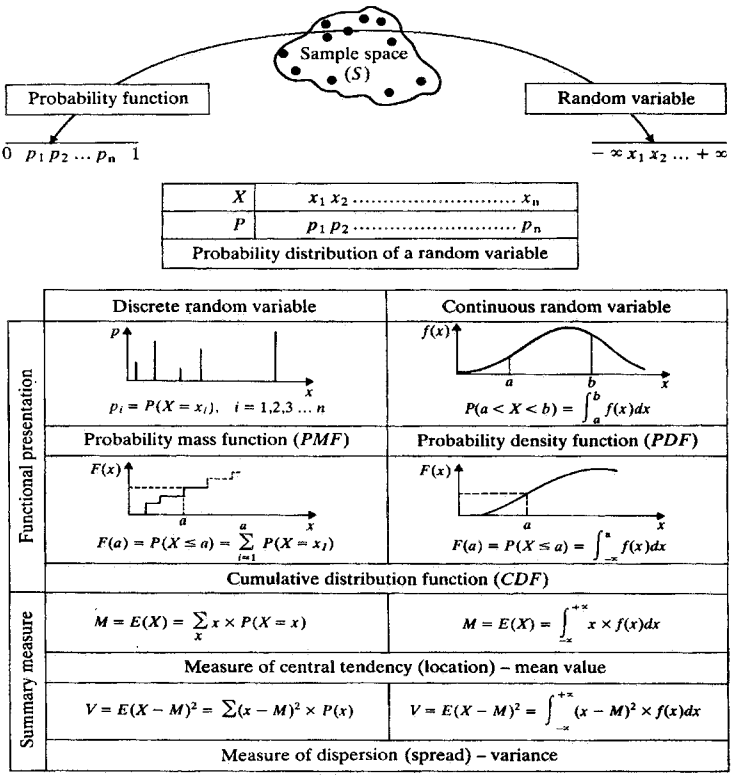

Figure 2-13. Probability System for Continuous Random Variable

The expectation of the deviation about the mean as a measure of variability, *E(X - M),* will not work because the expected deviation from the mean must be zero for obvious reasons. The solution is to find the *square* of each deviation from the mean, and then to find the expectation of the squared deviation. This characteristic is known as a *variance of the distribution, V,* thus:

$$V(X) = E(X - Mean)^2 = \sum (X - Mean)^2 \times P(x) \qquad \text{if X is discrete} \qquad (2.23)$$

$$V(X) = E(X - Mean)^2 = \int_{-\infty}^{+\infty} (X - Mean)^2 \times f(x)dx \quad \text{if X is continuous} \quad (2.24)$$

The positive square root of the variance of a distribution is called the *Standard Deviation, SD* (usually denoted using σ).

$$SD = \sigma = \sqrt{V(X)} \quad (2.25)$$

Probability distributions can be analyzed in greater depth by introducing other summary measures, known as *moments*. Very simply these are expectations of different powers of the random variable. More information about them can be found in texts on probability.

2.6.2.3 Variability

The standard deviation is a measure that shows how closely the values of random variables are concentrated around the mean. Sometimes it is difficult to use only knowledge of the standard deviation, to decide whether the dispersion is considerably large or small, because this will depend on the mean value. In this case the parameter known as coefficient of variation, CV_X, defined in equation (2.26) can be used.

$$CV_X = \frac{SD}{M} \quad (2.26)$$

Coefficient of variation is very useful because it gives better information regarding the dispersion. The concept thus discussed so far is summarized in Figure 2.13. In conclusion it can be said that the probability system is wholly abstract and axiomatic. Consequently, every fully defined probability problem has a unique solution.

2.7 DISCRETE PROBABILITY DISTRIBUTIONS

In probability theory, there are several rules that define the functional relationships between the possible values of random variable *X* and their probabilities, *P(X)*. As they are purely theoretical, i.e. they do not exist in reality, they are called *theoretical probability distributions*. Instead of analyzing the ways in which these rules have been derived, the analysis in this chapter concentrates on their properties. It is necessary to emphasize

that all theoretical distributions represent the family of distributions defined by a common rule through unspecified constants known as *parameters of distribution.* The particular member of the family is defined by fixing numerical values for the parameters, which define the distribution. The probability distributions most frequently used in reliability and Six Sigma are examined in this chapter.

Among the family of theoretical probability distributions that are related to discrete random variables, the Binomial distribution and the Poisson distribution are relevant to the objectives set by this book. A brief description of these two distributions is given below.

2.7.1 Bernoulli trials

The simple probability distribution is one with only two event classes. For example, a car is tested and one of two events, pass or fail, must occur, each with some probability. The type of experiment consisting of series of independent trials, each of which can eventuate in only one of two outcomes are known as *Bernoulli Trials*, and the two event classes and their associated probabilities a *Bernoulli Process.* In general, one of the two events is called a "success" and the other a "failure" or "nonsuccess". These names serve only to tell the events apart, and are not meant to bear any connotation of "goodness" of the event. The symbol p, stands for the probability of a success, q for the probability of failure $(p + q = 1)$. If 5 independent trials are made $(n = 5)$, then $2^5 = 32$ different sequences of possible outcomes (in general 2^n) would be observed.

The probability of given sequences depends upon p and q, the probability of the two events. Fortunately, since trials are independent, it is possible to compute the probability of any sequence.

If all possible sequences and their probabilities are written down the following fact emerges: *The probability of any given sequences of n independent Bernoulli Trials depends only on the number of successes and* p. This is regardless of the order in which successes and failure occur in a sequence. The corresponding probability is:

$$p^r q^{n-r}$$

Where r is the number of successes, and $n-r$ is the number of failures. Suppose that in a sequence of 10 trials, exactly 4 successes occur. Then the probability of that particular sequence is $p^4 q^6$. If $p = \frac{2}{3}$, then the probability can be worked out from:

$$\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^6$$

The same procedure is followed for any r successes out of n trials and for any p. Generalizing this idea for any r, n, and p, we have the following principle:

In sampling from the Bernoulli Process with the probability of a success equal to p, the probability of observing exactly r successes in n independent trials is:

$$P(r \text{ successes}|n,p) = \binom{n}{r} p^r q^{n-r} = \frac{n!}{r!(n-r)!} p^r q^{n-r} \qquad (2.27)$$

2.7.2 Binomial distribution

The theoretical probability distribution, which pairs the number of successes in n trials with its probability, is called the binominal distribution.

This probability distribution is related to experiments, which consist of a series of independent trials, each of which can result in only one of two outcomes: success or failure. By convention the symbol p stands for the probability of a success, q for the probability of failure $(p+q=1)$.

The number of successes, x in n trials is a discrete random variable which can take on only the whole values from 0 through n. The PMF of the Binomial distribution is given by:

$$PMF(x) = P(X=x) = \binom{n}{x} p^x q^{n-x}, \qquad 0 < x < n \qquad (2.28)$$

where:

$$\binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x} \qquad (2.29)$$

The binomial distribution expressed in cumulative form, representing the probability that X falls at or below a certain value 'a' is defined by the following equation:

$$P(X \leq a) = \sum_{i=o}^{a} P(X = x_i) = \sum_{i=0}^{a} \binom{n}{i} p^i q^{n-i} \tag{2.30}$$

$$E(X) = np \tag{2.31}$$

Similarly, because of the independence of trials, the variance of the binomial distribution is the sum of the variances of the individual trials, or $p(1-p)$ summed n times:

$$V(X) = np(1-p) = npq \tag{2.32}$$

Consequently, the standard deviation is equal to:

$$Sd(X) = \sqrt{npq} \tag{2.33}$$

As an illustration of the binomial distribution, the PMF and CDF are shown in Figure 2.14 with parameters $n = 10$ and $p = 0.3$.

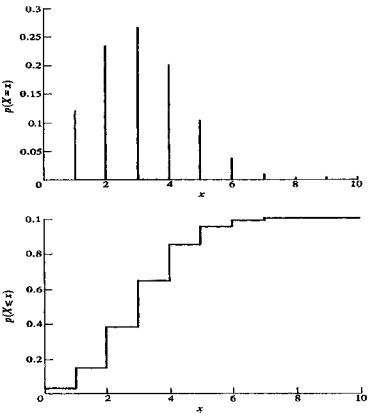

Figure 2-14. PMF and CDF For Binomial Distribution, n = 10, p = 0.3

Although the mathematical rule for the binomial distribution is the same regardless of the particular values which parameters n and p take, the shape of the probability mass function and the cumulative distribution function will depend upon them. The *PMF* of the binomial distribution is symmetric if $p = 0.5$, positively skewed if $p < 0.5$, and negatively skewed if $p > 0.5$.

2.7.3 Poisson distribution

The theoretical probability distribution which pairs the number of occurrences of an event in a given time period with its probability is called the Poisson distribution. There are experiments where it is not possible to observe a finite sequence of trials. Instead, observations take place over a continuum, such as time. For example, if the number of ions moving between plates in a battery in a given period of time is observed, say for one minute, it is difficult to think of this situation in terms of finite trials. If the number of binomial trials *n,* is made larger and larger and *p* smaller and smaller in such a way that *np* remains constant, then the probability distribution of the number of occurrences of the random variable approaches the Poisson distribution.

The probability mass function in the case of the Poisson distribution for random variable *X* can be expressed as follows:

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad \text{where } x = 0, 1, 2, \ldots \qquad (2.34)$$

λ is the *intensity of the process* and represents the expected number of occurrences in a time period of length *t*. Figure 2.15 shows the *PMF* of the Poisson distribution with $\lambda = 5$

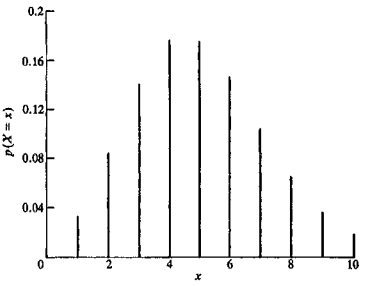

Figure 2-15. PMF of the Poisson distribution for $\lambda = 5$

The Cumulative Distribution Function for the Poisson distribution is given by:

$$F(x) = P(X \leq x) = \sum_{i=o}^{x} \frac{e^{-\lambda}\lambda^{i}}{i!} \tag{2.35}$$

The CDF of the Poisson distribution with $\lambda = 5$ is presented in Figure 2.16. Expected value of the distribution is given by

$$E(X) = \sum_{x=0} xP(X = x) = \sum_{x=0} x\frac{e^{-\lambda}\lambda^{x}}{x!}$$

Applying some simple mathematical transformations it can be proved that:

$$E(X) = \lambda \tag{2.36}$$

That is the expected number of occurrences in a period of time t is equal to λ. The variance of the Poisson distribution is equal to the mean:

$$V(X) = \lambda \tag{2.37}$$

Thus, the Poisson distribution is a single parameter distribution because it is completely defined by the parameter λ. In general, the Poisson distribution is positively skewed, although it is nearly symmetrical as λ becomes larger.

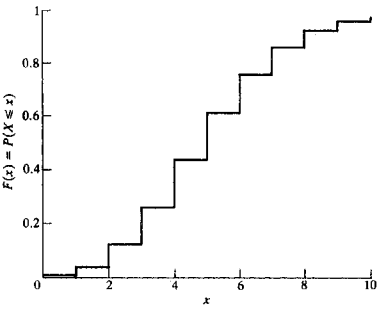

Figure 2-16. CDF of the Poisson distribution $\lambda = 5$

The Poisson distribution can be derived as a limiting form of the binomial if the following three assumptions were simultaneously satisfied:

1. n becomes large (that is, $n \rightarrow \infty$).
2. p becomes small (that is, $p \rightarrow 0$).

3. np remains constant.

Under these conditions, the binomial distribution with the parameters n and p, can be approximated by the Poisson distribution with parameter $\lambda = np$. This means that the Poisson distribution provides a good approximation to the binomial distribution if p is very small and n is large. Since p and q can be interchanged by simply interchanging the definitions of success and failure, the Poisson distribution is also a good approximation when p is close to one and n is large.

As an example of the use of the Poisson distribution as an approximation to the binomial distribution, the case in which $n = 10$ and $p = 0.10$ may be considered. The Poisson parameter for the approximation is then $\lambda = np = 10 \times 0.10 = 1$. The binomial distribution and the Poisson approximation are shown in Table 2.1.

The two distributions agree reasonably well. If more precision is desired, a possible rule of thumb is that the Poisson is a good approximation to the binomial if $n / p > 500$ (this should give accuracy to at least two decimal places).

Table 2-1. Poisson distribution as an approximation to the binomial distribution

x	Binomial $P(X = x\|n = 10, p = 0.1)$	Poisson $P(X = x\|\lambda = 1)$
0	0.598737	0.606531
1	0.315125	0.303265
2	0.074635	0.075816
3	0.010475	0.012636
4	0.000965	0.001580
5	0.000061	0.000158

2.8 CONTINUOUS PROBABILITY DISTRIBUTIONS

It is necessary to emphasize that all theoretical distributions represent the family of distributions defined by a common rule through unspecified constants known as *parameters of distribution.* The particular member of the family is defined by fixing numerical values for the parameters, which define the distribution. The probability distributions most frequently used in reliability and Six Sigma are examined in this chapter. Each of the above mentioned rules define a family of distribution functions. Each member of

the family is defined with a few parameters, which in their own way control the distribution. Parameters of a distribution can be classified in the following three categories (note that not all distributions will have all the three parameters, many distributions may have either one or two parameters):

1. *Scale parameter*, which controls the range of the distribution on the horizontal scale.
2. *Shape parameter*, which controls the shape of the distribution curves.
3. *Source parameter or Location parameter*, which defines the origin or the minimum value which random variable, can have. Location parameter also refers to the point on horizontal axis where the distribution is located.

Thus, individual members of a specific family of the probability distribution are defined by fixing numerical values for the above parameters.

2.8.1 Exponential distribution

Exponential distribution is fully defined by a single one parameter that governs the scale of the distribution. The probability density function of the exponential distribution is given by:

$$f(x) = \lambda \exp(-\lambda x), x > 0 \tag{2.38}$$

In Figure 2.17 several graphs are shown of exponential density functions with different values of λ. Notice that the exponential distribution is positively skewed, with the mode occurring at the smallest possible value, zero.

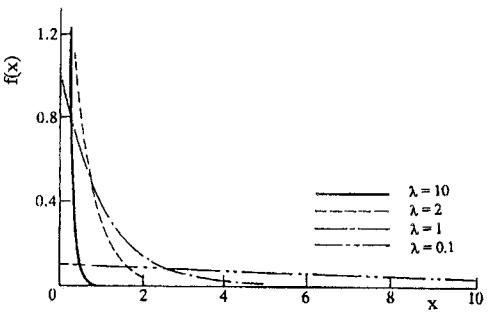

Figure 2-17. Probability density function of exponential distribution for different values of λ

The cumulative distribution of exponential distribution is given by:

$$F(x) = P(X < x) = 1 - \exp(-(\lambda x)) \tag{2.39}$$

It can be shown that the mean and variance of the exponential distribution are:

$$E(X) = 1/\lambda \tag{2.40}$$

$$V(X) = (1/\lambda)^2 \tag{2.41}$$

The standard deviation in the case of the exponential distribution rule has a numerical value identical to the mean, $SD(X) = E(X) = 1/\lambda$.

2.8.1.1 Memory less property

One of the unique properties of exponential distribution is that it is the only continuous distribution that has *memory less* property. Suppose that the random variable X measures the duration of time until the occurrence of failure of an item and that it is known that X has an exponential distribution with parameter λ. Suppose the present age of the item is t, that is X > t. Assume that we are interested in finding the probability that this item will not fail for another *s* units of time. This can be expressed using the conditional probability as:

$$P\{X > s+t | X > t\}$$

Using conditional probability of events, the above probability can be written as:

$$P\{X > s+t | X > t\} = \frac{P\{X > s+t \cap X > t\}}{P\{X > t\}} = \frac{P\{X > s+t\}}{P\{X > t\}} \tag{2.42}$$

However we know that for exponential distribution

$$P[X > s+t] = \exp(-\lambda(s+t)) \text{ and } P[X > t] = \exp(-\lambda t)$$

Substituting these expressions in equation (2.42), we get

$$P[X > s+t | X > t] = P[X > s] = \exp(-\lambda s)$$

That is, the conditional probability depends only on the remaining duration and is independent of the current age of the item. *This property is exploited to a great extent in reliability theory.*

2.8.2 Normal distribution

This is the most frequently used and most extensively covered theoretical distribution in the literature. The foundations of Six Sigma are based on the normal distribution. The Normal Distribution is continuous for all values of X between $-\infty$ and $+\infty$. It has a characteristic symmetrical shape, which means that the mean, the median and the mode have the same numerical value. The mathematical expression for its probability density function is as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right) \tag{2.43}$$

Where μ is the location parameter (as it locates the distribution on the horizontal axis), σ is the scale parameter (as it controls the range of the distribution). In a normal distribution, μ and σ also represents the mean and the standard deviation. The influence of the parameter μ on the location of the distribution on the horizontal axis is shown in Figure 2.18, where the values for parameter σ are constant.

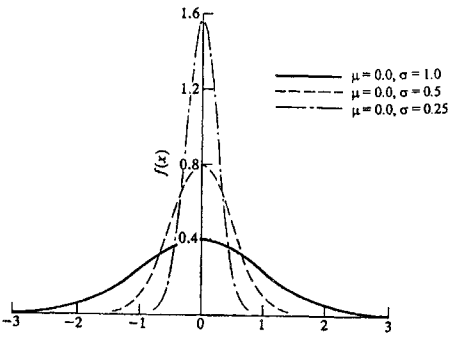

Figure 2-18. Probability density of normal distribution for different σ values

As the deviation of x from the location parameter μ is entered as a squared quantity, *two* different x values, showing the same absolute deviation from μ, will have the same probability density according to this rule. This dictates the symmetry of the normal distribution. Parameter μ can be any finite number, while σ can be any *positive* finite number. The cumulative distribution function for the normal distribution is:

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

Where $f(x)$ is the probability density function. Taking into account equation (2.43) this becomes:

$$F(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)dx \qquad (2.44)$$

In Figure 2.19 several cumulative distribution functions are given of the Normal Distribution, corresponding to different values of μ *and* σ. As the integral in equation (2.44) cannot be evaluated in a closed form, statisticians have constructed the table of probabilities, which comply with the normal rule for the standardized random variable, Z. Z is a theoretical random variable with parameters μ = 0 and σ = 1. The relationship between standardized random variable Z and random variable X is established by the following expression:

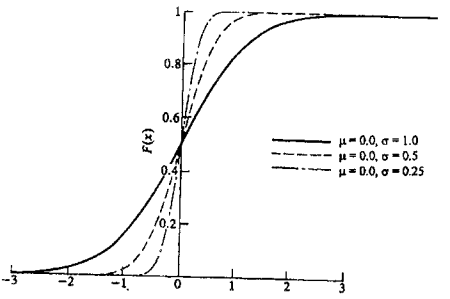

Figure 2-19. Cumulative distribution of normal distribution for different values of μ and σ

$$z = \frac{x - \mu}{\sigma} \tag{2.45}$$

Making use of the above expression the equation (2.43) becomes simpler:

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \tag{2.46}$$

The standardized form of the distribution makes it possible to use only one table for the determination of *PDF* for any normal distribution, regardless of its particular parameters (see Table in appendix). The relationship between $f(x)$ and $f(z)$ is :

$$f(x) = \frac{f(z)}{\sigma} \tag{2.47}$$

By substituting $\frac{x - \mu}{\sigma}$ in place of z, equation. (2.44) becomes:

$$F(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz = \Phi\left(\frac{a - \mu}{\sigma}\right) \tag{2.48}$$

Where Φ is the standard normal distribution function and is given by:

$$\Phi(z) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dx \tag{2.49}$$

The corresponding standard normal probability density function is:

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{2.50}$$

Most tables of the normal distribution give the cumulative probabilities for various *standardized* values. That is, for a given z value the table

provides the cumulative probability up to and including that standardized value in a normal distribution. In *Microsoft EXCEL®*, the cumulative distribution function and density function of normal distribution with mean μ and standard deviation σ can be found using the following function.

F(x) = NORMDIST (x, μ, σ, TRUE), and f(x) = NORMDIST (x, μ, σ, FALSE)

If F(z) value is known, then to find the value of z using *Microsoft EXCEL* the following function may be used.

z = NORMINV(F(z), 0, 1)

Note that the value of z is nothing but the Sigma level quality as measured in Six Sigma measurement system (see chapter 3 for more details). The expected value of the normal random variable is equal to the location parameter, μ. That is:

$$E(X) = \mu \tag{2.51}$$

Whereas the variance is

$$V(X) = \sigma^2 \tag{2.52}$$

Since normal distribution is symmetrical about its mean, the area between μ - kσ, μ + kσ (k is any real number) takes a unique value, which is shown in Figure 2.20

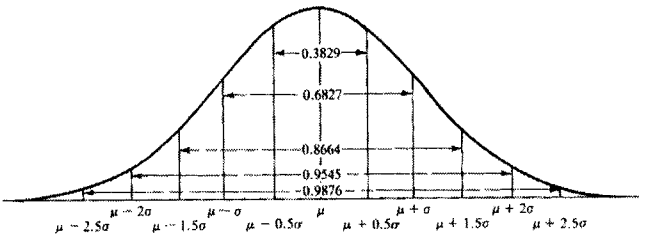

Figure 2-20. The areas under a normal distribution between μ- kσ and μ+kσ

Table 2.2 gives the area between μ - kσ, μ + kσ for k = 1, 2, ..., 6.

Table 2-2. Area between $\mu - k\sigma$, $\mu + k\sigma$ under a normal distribution

Range	Area (cumulative probability)
$\mu - 1\ \sigma$ and $\mu + 1\ \sigma$	0.68268948
$\mu - 2\ \sigma$ and $\mu + 2\sigma$	0.954499876
$\mu - 3\sigma$ and $\mu + 3\sigma$	0.997300066
$\mu - 4\sigma$ and $\mu + 4\sigma$	0.999936628
$-\infty$ and $\mu + 4.5\ \sigma$	0.999996599
$\mu - 5\sigma$ and $\mu + 5\sigma$	0.999999426
$\mu - 6\sigma$ and $\mu + 6\sigma$	0.999999998

The area between $-\infty$ and $\mu + 4.5\ \sigma$ gives the defect rate of 3.4 out of one million which is the target for Six Sigma quality.

2.8.2.1 Central limit theorem

Suppose $X_1, X_2, \ldots X_n$ are mutually independent observations of a random variable X having a well-defined mean μ_x and standard deviation σ_x. Let

$$Z_n = \frac{\bar{X} - \mu_x}{\sigma_x / n} \tag{2.53}$$

Where,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2.54}$$

and $F_{z_n}(z)$ be the cumulative distribution function of the random variable Z_n. Then for all z, $-\infty < z < \infty$,

$$\lim_{n \to \infty} F_{Z_n}(z) = F_Z(z) \tag{2.55}$$

Where $F_Z(z)$ is the cumulative distribution of standard normal distribution *N(0,1)*. The X values have to be from the same distribution but the remarkable feature is that this distribution does not have to be normal, it can be uniform, exponential, beta, gamma, Weibull or even an unknown one.

2.8.3 Lognormal distribution

The lognormal probability distribution, can in some respect, be considered as a special case of the normal distribution because of the derivation of its probability function. If a random variable $Y = \ln X$ is normally distributed then, the random variable X follows the lognormal distribution. Thus, the probability density function of a log-normal random variable X is defined as:

$$f_X(x) = \frac{1}{x\sigma_l\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu_l}{\sigma_l} \right)^2 \right) \quad \geq 0 \tag{2.56}$$

The parameter μ_l is called the *scale parameter* (see Figure 2.21) and parameter σ_l is called the *shape parameter*. The relationship between parameters μ (location parameter of the normal distribution) and μ_l is defined by the following expression:

$$\mu = \exp\left(\mu_l + \frac{1}{2}\sigma_l^2 \right) \tag{2.57}$$

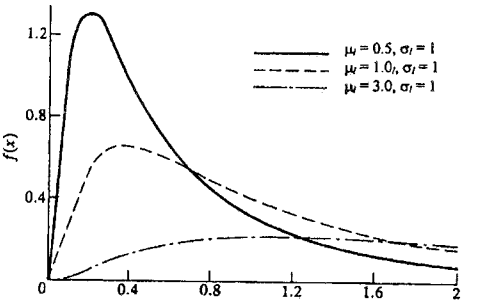

Figure 2-21. Probability density of log-normal distribution

The cumulative distribution function for the lognormal distribution is defined by the following expression:

$$F_X(x) = P(X \le x) = \int_0^x \frac{1}{x\sigma_l\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu_l}{\sigma_l}\right)^2\right) dx \quad (2.58)$$

As the integral cannot be evaluated in closed form, the same procedure is applied as in the case of normal distribution to calculate the cumulative distribution of a log-normal distribution. Thus, making use of the standardized random variable, equation (2.58) transforms into:

$$F_X(x) = P(X \le x) = \Phi\left(\frac{\ln x - \mu_l}{\sigma_l}\right) \quad (2.59)$$

The measures of central tendency in the case of lognormal distributions are defined by the:

(a) Location parameter (Mean)

$$M = E(X) = \exp\left(\mu_l + \frac{1}{2}\sigma_l^2\right) \quad (2.60)$$

(b) Deviation parameter (the variance)

$$V(X) = \exp\left(2\mu_l + \sigma_l\right)^2 \left[\exp(\sigma_l^2 - 1)\right] \quad (2.61)$$

2.8.4 Weibull distribution

This distribution originated from the experimentally observed variations in the yield strength of Bofors steel, the size distribution of fly ash, fiber strength of Indian cotton, and the fatigue life of a *St*-37 steel by the Swedish engineer W.Weibull. As the Weibull distribution has no characteristic shape, such as the normal distribution, it has a very important role in the statistical analysis of experimental data. The shape of this distribution is governed by its parameter. The rule for the probability density function of the Weibull distribution is:

$$f(x)=\frac{\beta}{\eta}\left(\frac{x-\gamma}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{x-\gamma}{\eta}\right)^{\beta}\right], \quad \eta,\beta,\gamma\geq 0 \quad x\geq\gamma \quad (2.62)$$

where η, β, $\gamma > 0$. As the location parameter γ is often set equal to zero, in such cases:

$$f(x)=\frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{x}{\eta}\right)^{\beta}\right]$$

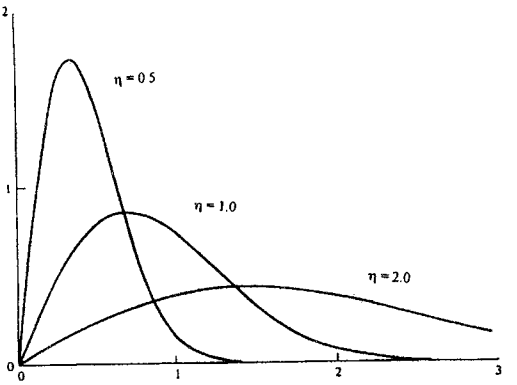

Figure 2-22. Probability density of Weibull distribution with $\beta = 2.0$, $\gamma = 0$, $\eta = 0.5, 1, 2$

For different parameter values of β, the Weibull distribution takes different shapes. For example, when $\beta = 3.4$ the Weibull approximates to the normal distribution; when $\beta=1$, it is identical to the exponential distribution. Figure 2.22 shows the Weibull probability density function for selected parameter values. The cumulative distribution functions for the Weibull distribution is:

$$F(x)=1-\exp\left[-\left(\frac{x-\gamma}{\eta}\right)^{B}\right] \quad (2.63)$$

For $\gamma = 0$, the cumulative distribution is given by

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\eta}\right)^{\beta}\right] \tag{2.64}$$

The expected value of the Weibull distribution is given by:

$$E(X) = \gamma + \eta \times \Gamma\left(\frac{1}{\beta} + 1\right) \tag{2.65}$$

where Γ is the gamma function, defined as

$$\Gamma(n) = \int_0^{\infty} e^{-x} \times x^{n-1} dx$$

When n is integer then $\Gamma(n) = (n-1)!$. For other values, one has to solve the above integral to the value. Values for this can be found in Gamma function table given in the appendix. In *Microsoft EXCEL,* Gamma function, $\Gamma(x)$ can be found using the function, *EXP[GAMMALN(x)].* The variance of the Weibull distribution is given by:

$$V(X) = (\eta)^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right] \tag{2.66}$$

2.9 STOCHASTIC PROCESSES

Stochastic process (also known as random process) is a collection of random variables $\{X(t), t \in T\}$, where T is the set of numbers that indexes the random variables X(t). In reliability, it is often appropriate to interpret t as time and T as the range of time being considered. The set of possible values the stochastic process X(t) can assume is called state. The set of possible states constitutes the state-space, denoted by E. The state-space can be continuous or discrete. For example consider a system with two items connected in parallel. Assume that the time-to-failure distributions of the two parallel items are given by two sequences of random variables X_i and Y_i (i = 1, 2 ...). Here the subscript i represents the time to i^{th} failure of the items. If the sequence of random variable Z_i represents the i^{th} repair time, then the process { X(t), $t \geq 0$ } by definition forms a stochastic process. At any time t,

it is possible that two, one or none of these two items will be maintaining the required function. Thus, the set {0, 1, 2} forms the state-space of the system.

Analyzing a system using stochastic processes will involve the following fundamental steps.

1. Identify the time domain T for the system. The time domain T can be discrete or continuous.
2. Identify the state space of the system. The state space can be either discrete or continuous.

Once the process is defined using the family of random variables {X(t), t ∈ T}, state space (E) and the parameter set (T), the next step will be to identify the properties of the process that can be used to classify the process and also to analyze the process to extract information. As far as reliability is concerned, processes with a continuous time parameter and discrete state space are important. In this chapter, we discuss the following stochastic processes.

1. Markov processes
2. Non-homogeneous Poisson Processes
3. Renewal processes

Readers who are interested to know more on applications of stochastic process are advised to refer to Birolini *(1997).*

2.10 MARKOV PROCESSES

A stochastic process is said to be a Markov process if the future evolution of the process depends only on the current time and state. That is, the future state of a system is conditionally independent of the past, given that the present state and age of the system is known. Thus, to predict the future state one need to know only the present state and age of the system. Mathematically, a stochastic process {X(t); t ∈ T} with state-space E is called a Markov process if it satisfies the condition:

$$\begin{aligned} P[X(t_n + h) = j \mid X(t_n) = i_n, X(t_{n-1}) = i_{n-1}, \ldots, X(t_0) = i_0] \\ = P[X(t_n + h) = j \mid X(t_n) = i_n] \end{aligned} \tag{2.67}$$

for all $(j, i_n, i_{n-1}, \ldots, i_0) \in E$. The above property is called *Markov property.* A Markov process with discrete state space is called *Markov Chain.* A Markov process with continuous time and discrete state space is called *continuous time Markov chain (CTMC).* The conditional probability

defined in equation (2.67) is referred as the *transition probability* of Markov process and is defined using the notation $P_{ij}(t_n + h)$

$$P_{ij}(t_n + h) = P[X(t_n + h) = j \mid X(t_n) = i_n] \tag{2.68}$$

A Markov process is called *time-homogeneous* or *stationary* if the transition probabilities are independent of time t. For a stationary Markov process,

$$P_{ij}(t_n + h) = P_{ij}(t_0 + h) = P_{ij}(h) \tag{2.69}$$

Thus, the transition from state i to state j in a stationary Markov chain depends only on the duration h. The transition probabilities $P_{ij}(t + h)$ satisfy the following *Chapman-Kolmogrov* equations

$$P_{ij}(t + h) = \sum_{k \in E} P_{ik}(h) P_{kj}(t) \tag{2.70}$$

In all the models discussed in this Chapter we assume that the Markov process is stationary. It is convenient to use a matrix to represent various state transition probabilities of a Markov process. For example, if a system has n states, we define a matrix **P**, such that

$$\mathbf{P} = [P_{ij}(h)] = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ . & . & . & . \\ . & . & . & . \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

The matrix **P** is called *Transition Probability Matrix* (TPM) or *Stochastic Matrix.*

Let $\{S_j, j \in E\}$ represent the time spent at state *j* (sojourn time at state *j*). The probability that the process will spend more than *t* hours at state *j* is, $P[S_j > t]$. Assume that the process has already spent *h* hours in state *j*, the probability that it will spend additional *t* hours in state *j* is given by:

$$P[S_j > t + h \mid S_j > h] \tag{2.71}$$

Since past is irrelevant in Markov process, the above expression can be written as:

$$P[S_j > t + h \mid S_j > h] = P[S_j > t] \tag{2.72}$$

The only continuous distribution that satisfies the above relation is exponential distribution. The above property of exponential distribution is called *memory-less property.* Thus, in a Markov process, the time spent in any state follows exponential distribution. Thus,

$$P[S_j > t] = \exp(-\nu_j t) \tag{2.73}$$

where the parameter ν_i depends on state *i*. This is a very important result and limitation of Markov processes. This implies that the Markov process can be applied in reliability theory only when the time-to-failure follow exponential distribution.

Transition Rates between the States of a Markov Process

Since the time spent at any state *j* of a Markov process follows exponential distribution, the probability that the process remains in state *j* during a small interval δt is given by:

$$\begin{aligned} P[S_j > \delta t] &= \exp(-\nu_j \delta t) \\ &= 1 - \frac{\nu_j \delta t}{1!} + \frac{(\nu_j \delta t)^2}{2!} - \ldots \\ &= 1 - \nu_j \delta t + O(\delta t) \end{aligned}$$

where $O(\delta t)$ represents the terms which are negligible as δ approaches zero. That is,

$$\lim_{\delta t \to 0} \frac{O(\delta t)}{\delta t} = 0$$

Thus, for a small duration of δt, $P_{jj}(\delta t)$, probability that the process will remain in state *j* for small duration δt is given by:

$$P_{jj}(\delta t) = 1 - \nu_j \delta t + O(\delta t)$$

Probability that the system will leave state *j* is given by

$$1 - P_{jj}(\delta t) = \nu_j \delta t + O(\delta t)$$

ν_j is the rate at which the process {X(t), t ∈ T} leaves the state *j*. Rearranging the above equation we have,

$$P_{jj}(\delta t) - 1 = -\nu_j \delta t + O(\delta t)$$

Substituting $\lambda_{jj} = -\nu_j$ in the above equation, we get

$$P_{jj}(\delta t) - 1 = \lambda_{ii} \delta t + O(\delta t)$$

It is easy to verify that

$$\lim_{\delta t \to 0} \frac{P_{jj} - 1}{\delta t} = \lambda_{jj} \tag{2.74}$$

The transition probability $P_{ij}(\delta t)$, that is the process will enter state *j* (with probability r_{ij}) after leaving state *i* during a small duration δt is given by:

$$\begin{aligned} P_{ij}(\delta t) &= [1 - P_{ii}(\delta t)] \times r_{ij} = [\nu_i \delta t + O(\delta t)] \times r_{ij} \\ &= \lambda_{ij} \delta t + O(\delta t) \end{aligned} \tag{2.75}$$

where λ_{ij} is the rate at which the process enters the state *j* from the state *i*.

Let $P_j(t) = P[X(t) = j]$, that is $P_j(t)$ denotes that the process is in state *j* at time t. Now for any δt, we have

$$\begin{aligned} P_j(t + \delta t) &= P[X(t + \delta t) = j] \\ &= \sum_{i \in E} P[X(t + \delta t) = j \mid X(t) = i] P[X(t) = i] \end{aligned}$$

The above expression can be written as

$$P_j(t + \delta t) = \sum_{i \in E} P_{ij}(\delta t) P_{i(t)} \tag{2.76}$$

The above equation (2.76), upon few mathematical manipulation will give a system of differential equations which can be solved to find $P_i(t)$.

From equation (2.76)

$$P_j(t+\delta t) - P_j(t) = \sum_{\substack{i \in E \\ i \neq j}} P_{ij}(\delta t) P_i(t) + P_j(t)[P_{jj}(\delta t) - 1] \tag{2.77}$$

For $\delta t \rightarrow 0$, and using equation (2.74) and (2.75), equation (2.77) can be written as:

$$\frac{d}{dt} P_j(t) = \sum_{i \in E} \lambda_{ij} P_i(t) = \sum_{\substack{i \in E \\ i \neq J}} \lambda_{ij} P_i(t) - \nu_j P_j(t) \tag{2.78}$$

Also

$$\sum_{j \in E} P_j(t) = 1 \tag{2.79}$$

Equation (2.78) is called *Kolmogrov backward equations*, which along with equation (2.79) has a unique solution. Thus, various state probabilities of the process can be obtained by solving the system of differential equations of the form:

$$\frac{d}{dt} P(t) = \Delta P(t) \tag{2.80}$$

where P(t) is a time-dependent N dimensional probability vector and Δ is a square matrix where the element (i,j) represents the rate at which the process enters the state *j* from the state *i*.

2.11 NON-HOMOGENEOUS POISSON PROCESS

A counting process $\{N(t),\ t \geq 0\}$ is said to be a *non-homogeneous Poisson process* with intensity function $\lambda(t),\ t \geq 0$, if:

1. $N(0) = 0$.

2. *N(t)* has independent increments.
3. The number of events in any interval *t* and *t + s* has a Poisson distribution with mean [*S(t+s) - S(t)*], that is

$$P[N(t+s)-N(t)=n]=\frac{[S(t+s)-S(t)]^n \exp\{-(S(t+s)-S(t))\}}{n!} \tag{2.81}$$

Where

$$S(t)=\int_0^t \lambda(x)dx \tag{2.82}$$

S(t) is the expected number of events in (0,t). Also, *N(t+s) - N(t)* is Poisson distributed with mean *S(t+s) - S(t).*

2.12 RENEWAL PROCESS

Renewal theory was originally used to analyze the replacement of equipment upon failure, to find the distribution of number of replacement and mean number of replacements. Let $\{X_n; n = 1, 2, \ldots\}$ be a sequence of non-negative independent random variables with common distribution F. Let X_n be the time between $(n-1)^{st}$ and n^{th} event. Let:

$$S_0 = 0,\ S_n = \sum_{i=1}^{n} X_i \tag{2.83}$$

Thus S_n is the time to n^{th} event or *epoch* at which the nth renewal occurs. Let N(t) be the number of renewals by time t.

$$N(t) = Max\{n; \leftrightarrow \quad S_n \leq t\} \tag{2.84}$$

Let X_1, X_2,, ... are independent and identically distributed random variables with distribution *F(t)*. Then $P\{S_n \leq t\}$ is given by:

$$P\{S_n(t) \leq t\} = F^n(t) \tag{2.85}$$

where $F^n(t)$ is the n-fold convolution of $F(t)$. That is,

$$F^n(t) = \int_0^t F^{n-1}(x)dF(x) \tag{2.86}$$

We use the convention that $F^0(t) = 1$ for $t > 0$. $F^n(t)$ represents the probability that the nth renewal occurs by time t. The distribution of N(t) can be derived using the following arguments.

Distribution of N(t)

The counting process, N(t), is called a renewal process. From the definition of N(t) and S_n, we have

$$\{N(t) = n\} \Leftrightarrow \{S_n \leq t, S_{n+1} > t\} \tag{2.87}$$

$$\begin{aligned} P[N(t) = n] &= P\{N(t) < n+1\} - P\{N(t) < n\} \\ &= P\{S_{n+1} > t\} - P\{S_n > t\} \\ &= 1 - F^{n+1}(t) - [1 - F^n(t)] \end{aligned} \tag{2.88}$$

Thus the probability that the number of renewal by time t is equal to n, is given by:

$$P\{N(t) = n\} = F^n(t) - F^{n+1}(t) \tag{2.89}$$

It is difficult to evaluate the above function analytically for many theoretical distributions, however it can be solved using well-known numerical methods.

2.12.1 Renewal Function

The expected number of renewals during specified duration t is given by:

$$E[N(t)] = M(t) = \sum_{i=1}^{\infty} i \times [F^i(t) - F^{i+1}(t)] \tag{2.90}$$

The above equation can be simplified, and the expected number of renewals (expected number of demands) is given by:

$$M(t) = \sum_{i=1}^{\infty} F^i(t) \tag{2.91}$$

The above equation is called *renewal function, M(t),* and it gives the number of renewals during (0, t]. Taking the derivative of renewal function we get:

$$m(t) = \frac{d}{dt} M(t) = \sum_{n=1}^{\infty} f^n(t) \tag{2.92}$$

Where $f^n(t)$ is the derivative of $F^n(t)$. $m(t)\delta t$ is the probability that a renewal occurs during (t, t+δt). *m(t)* is called the *renewal density* or *renewal rate.*

Calculating $F^n(t)$, P[N(t) = n], M(t) and m(t)

Exponential Distribution

$$F(t) = 1 - \exp(-\lambda t)$$

When the time to failure distribution is exponential, the renewal process constitutes a Poisson process. Thus, Poisson process is also a special case of renewal process where time to failure is exponential.

$$F^n(t) = 1 - \sum_{i=0}^{n-1} \frac{\exp(-\lambda t) \times (\lambda t)^i}{i!} \tag{2.93}$$

$$P[N(t) = n] = \frac{\exp(-\lambda t) \times (\lambda t)^n}{n!} \tag{2.94}$$

$$M(t) = \lambda t \tag{2.95}$$

$$m(t) = \lambda \tag{2.96}$$

Normal Distribution

By assuming $\sigma << \mu$, we have

$F^n(t) = \Phi(\frac{t - n \times \mu}{\sigma \times \sqrt{n}})$, where $\Phi(t)$ is the standard normal distribution.

The distribution of N(t) is given by:

$$P[N(t) = n] = \sum_{n=1}^{\infty} [\Phi(\frac{t - n \times \mu}{\sigma \times \sqrt{n}}) - \Phi(\frac{t - (n+1) \times \mu}{\sigma \times \sqrt{n+1}}]$$

$$M(t) = \sum_{n=1}^{\infty} \Phi(\frac{t - n \times \mu}{\sigma \times \sqrt{n}}) \tag{2.97}$$

For distributions like Weibull, one has to use numerical approximation to find the renewal function.

2.12.2 Elementary Renewal Theorem

For a distribution function F(t) with F(0) = 0 and finite mean, and if f(x) exists then the following equation is valid

$$\lim_{t \to \infty} \frac{M(t)}{t} = \frac{1}{MTTF} \tag{2.98}$$

The above result is called the *Elementary Renewal Theorem.* This implies that in the steady state, the expected number of failures is given by the ratio of t over the MTTF value.

http://www.springer.com/978-0-387-30255-3

Reliability and Six Sigma

Kumar, U.D.; Crocker, J.; Chitra, T.; Saranga, H.

2006, XX, 386 p. 81 illus., Hardcover
ISBN: 978-0-387-30255-3