CHAPTER 2

The Preliminaries

Introduction

An appreciation of certain basic probability concepts is essential for the development of reliability models and their subsequent evaluation. In general, probability mathematics provide the medium for examination of systems which exhibit random phenomena, i.e. behave in accordance with probablistic rather than deterministic laws. It has been assumed that the reader is familiar with the basic probability concepts normally encountered in an undergraduate engineering course. Many text books and, therefore, courses introduce probability theory either as an abstract mathematical concept or through the use of *apriori* situations such as dice or playing-card type examples. In these cases, it is often difficult to develop an easy and effective interface between the basic reliability and probability concepts. This chapter reviews the basic probability theory required in subsequent chapters with emphasis on utilization in system reliability modelling and evaluation.

Sample Space

The set of all possible outcomes of a random phenomenon is called the sample space, sample description space or possibility space. As an example consider two transmission links each existing either in the up state (U) or the down state (D). The description of the possible states at any time is given by the set

$$S = \{(1U, 2U), (1U, 2D), (1D, 2U), (1D, 2D)\}$$

A set which has a definite number of elements is called a finite set. A set which is not finite but whose elements can be put in a one to one correspondence with the set of natural numbers is said to be countably infinite or denumerable. Both of these types, the finite and the denumerable set come under the general name of countable set.

As another example, consider the load on a pumping system. This may assume any value between L_0, the minimum load and L_1 the maximum peak load. The sample space S, therefore, consists of all points s such that

$$L_0 \leqslant s \leqslant L_1$$

The interval (L_0, L_1) contains a noncountable infinity of members. Such a sample space which is not countable is called uncountably infinite or simply uncountable.

Events

Consider the descriptions $(1U, 2D)$ and $(1D, 2U)$ in the example of two transmission links. These descriptions define the event that one of the transmission links has failed. Similarly if both the transmission links are needed to keep the system in operation, the set $\{(1D, 2D), (1U, 2D), (1D, 2U)\}$ defines the event that the system has failed. An event can therefore be defined as a set of descriptions and the event E is said to have occurred if the outcome of the random phenomenon is a member of E. As the sample space contains the descriptions of all possible outcomes of the random phenomenon under consideration, an event may also be defined as any subset of the sample description space. In the case of demand on the pumping system, the subset $A = \{s: L_2 \leqslant s \leqslant L_1\}$ defines the event that load is greater than or equal to L_2. Whenever the load is in the interval (L_2, L_1), the event 'load is greater than L_2' is said to have occurred. The events are sets and therefore the algebra of events is essentially the concepts of set theory.

Random Variables

A random variable is a quantity which assumes values in accordance with certain probabilistic laws. A random variable which assumes discrete values is called a discrete random variable and one which assumes values from a continuous interval is termed a continuous random variable. This definition of a random variable is sufficient for the purpose of this book. A more precise definition of a random variable is as a function defined on a sample space in which certain technical conditions are satisfied. The random variable relabels the descriptions of outcomes contained in S in terms of real numbers. The domain of the random variable is the sample space S and the range is contained in the set Re of all real numbers.

Consider again the example of the two transmission links. Let X be the function defined on the sample space S, where X denotes the number of components down. The values of this discrete random variable are given below:

$$\begin{aligned} X(s) &= 0 \ , \ s = (1U, 2U) \\ &= 1 \ , \ s = (1U, 2D), (1D, 2U) \\ &= 2 \ , \ s = (1D, 2D) \end{aligned}$$

In the above example, instead of describing the outcomes as shown above, they could be assigned integer values by a random variable Z such that

$$\begin{aligned} Z &= 0 \ , \ s = (1U, 2U) \\ &= 1 \ , \ s = (1U, 2D) \\ &= 2 \ , \ s = (1D, 2U) \\ &= 3 \ , \ s = (1D, 2D) \end{aligned}$$

The random variable Z depicts the state of the system and its different values are termed the state space. If a device is put into operation at time $t = 0$, the time to failure can be denoted by the continuous random variable X. Some other examples of a continuous random variable X assuming a value x are temperature $(-273^0\text{C} \leqslant x < \infty)$, the load on a power system $(L_0 \leqslant x \leqslant L_1)$, the repair time of a component $(0 < x < \infty)$ and the noise voltage at an amplifier output point.

Probability Laws

In application, the functional form of the random variable is not usually of great interest. The main focus is usually on the probability with which the random variable assumes a certain value. Probability is a function which assigns a number between 0 and 1 to a set of points (event) in S. The statement that the random variable X assumes a value in set B of real numbers, implies that the event defined by the subset, $(s: X(s)$ is in $B)$ occurs. Therefore

$$P[\{s: X(s) \text{ is in } B\}] = P[X \text{ is in } B]$$

This is the basic formula for obtaining the probability function of the random variable X from the probability function which exists on the sample space S on which the random variable X is defined as a function. Some other types of events in terms of a random variable can be defined for fixed numbers x, a, b

$$\begin{aligned} [X = x] &= \{s: X(s) = x\} \\ [X \leqslant x] &= \{s: X(s) \leqslant x\} \\ [X > x] &= \{s: X(s) > x\} \\ [a < X < b] &= \{s: a < X(s) < b\} \end{aligned}$$

A discrete random variable assumes only discrete values $x_i, i = 0, 1, 2, \ldots$ from

the set Re of real numbers. The probability density function for a discrete random variable is defined by

$$p_X(x) = P[X = x] \tag{2.1}$$

This function should clearly have the following properties

(i) $p_X(x) = 0$ unless x is one of $x_0, x_1, x_2, \ldots$

(ii) $0 \leqslant p_X(x_i) \leqslant 1$

(iii) $\sum_i p_X(x) = \sum_i P[X = x_i] = 1$

The probability density function for a discrete random variable is sometimes also called the probability mass function of X. The probability function $P_X(B)$ of the random variable in terms of probability mass function is given by

$$P_X(B) = P[X \text{ is in } B] = \sum_{x_i \in B} p_X(x_i)$$

The probability distribution function $F_X(x)$ of the discrete random variable is given by

$$F_X(x) = P[X \leqslant x]$$

$$= \sum_{x_i \leqslant x} p_X(x_i)$$

It should be noted that the domain of the distribution function is the set of all real numbers and the range, being a probability is the interval (0,1). This is in contrast to the case of a random variable whose domain is the sample space.

A continuous random variable can assume any value over a continuous interval. Since the number of elementary events in a continuous interval is infinite, the probability of the random variable X assuming a value exactly equal to x is zero. This is, however, not an impossible event. For example, although the time to actual failure of a component may be x, the probability of this event happening is zero. The probability density function of the form in Equation (2.1) is therefore not suitable for a continuous random variable. The probability density function $f_X(x)$ of a continuous random variable X is so defined that

$$P[a \leqslant X \leqslant b] = \int_a^b f_X(y)\,dy \tag{2.2}$$

The probability distribution function $F_X(x)$ can be now written as

$$F_X(x) = P[-\infty < X \leqslant x]$$

$$= \int_{-\infty}^{x} f_X(y)\,dy \tag{2.3}$$

It follows from Equation (2.3) that

$$f_X(x) = F'_X(x) \tag{2.4}$$

The function $f_X(x)$ has the following properties

1. $f_X(x)$ is non-negative

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

3. The function $f_X(x)$ is continuous at all but a finite number of points, i.e. it is piece-wise continuous.

When we are dealing with the distributions of operating and down times, the random variable X is non-negative and its density function is zero over the negative range. It is sometimes more convenient to work with the complementary function of $F(x)$ called the survivor function

$$\mathfrak{F}(x) = P[X > x]$$

$$= \int_x^{\infty} f(y)\,dy \tag{2.5}$$

$$= 1 - F(x)$$

It follows from this expression that

$$f(x) = -\mathfrak{F}'(x) \tag{2.6}$$

In reliability modelling and evaluation one function which is used extensively and which is equivalent to $f(x)$ is the hazard function. In practice, depending upon the circumstances employed, it may be known by a variety of names, age specific failure rate or simply failure rate, repair rate, hazard rate, force of mortality etc. A detailed interpretation of this function is given in the next

chapter while developing the concept of frequency. This can be defined as

$$\phi(x) = \lim_{\Delta x \to 0^+} \frac{P[x < X \leqslant x + \Delta x | x < X]}{\Delta x} \tag{2.7}$$

With the interval length approaching zero, Equation (2.2) can be written as

$$f_X(x) = \lim_{\Delta x \to 0^+} \frac{P[x < X \leqslant x + \Delta x]}{\Delta x}.$$

Equation (2.7) can be rewritten in the form

$$\phi(x) = \lim_{\Delta x \to 0^+} \frac{P[x < X \leqslant x + \Delta x]}{\Delta x} \quad \frac{1}{P[x < X]} \tag{2.8}$$

$$= \frac{f_X(x)}{\mathfrak{F}_X(x)}$$

It is usual to drop the suffix X denoting the random variable unless there is a likelihood of incorrect interpretation and this has been done in this text. Using Equations (2.6) and (2.8)

$$\phi(x) = -\frac{\mathfrak{F}'(x)}{\mathfrak{F}(x)}$$

$$= -\frac{d}{dx}[\log \mathfrak{F}(x)] \tag{2.9}$$

Integrating Equation (2.9) and using the condition that $\mathfrak{F}(0) = 1$

$$\mathfrak{F}(x) = \exp\left[-\int_0^x \phi(y)\, dy\right]$$

and therefore

$$f(x) = \phi(x) \exp\left[-\int_0^x \phi(y)\, dy\right]$$

This expression shows that $\phi(x)$, the hazard function uniquely determines the probability density function. The probability density function, the probability distribution function, the survivor function and the hazard rate function are mathematically equivalent.

Expectation

The probabilistic behaviour of a random variable is completely defined by the probability density function or distribution function. It is, however, often of interest to obtain a single value which may represent the random variable and its probability distribution. One such characteristic value is the expectation or mean. This is denoted by $E(x)$ and given by

$$E(X) = \sum_i x_i P[X = x_i]$$

$$= \sum_i x_i p(x_i) \quad \text{if } X \text{ is a discrete random variable}$$

or

$$= \int_{-\infty}^{+\infty} x f(x)\, dx \quad \text{if } X \text{ is a continuous random variable} \tag{2.10}$$

The expectation is said to exist if the series or the integral involved converges absolutely, i.e.

$$\sum_i |x_i|\, p(x_i) < \infty \quad \text{for the discrete case}$$

and

$$\int_{-\infty}^{+\infty} |x|\, f(x)\, dx < \infty \quad \text{for the continuous case}$$

The expectation or the mean value has a meaningful interpretation in terms of the average of a sample. This interpretation is provided by the law of large numbers. Assume that there are n random variables $X_1, X_2, \ldots, X_n$ which are identically distributed as X and each has a mean m. This set of n random variables represents a random sample of n observations $X_1, X_2, \ldots, X_n$. The sample mean, which is also a random variable is given by

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

According to the law of large numbers for any constant $c > 0$

$$\lim_{n \to \infty} P[|\bar{X} - m| > c] = 0 \tag{2.11}$$

This implies that as the sample size increases, the sample mean approaches the mean of the random variable. Thus if the random variable X is observed many times and each time the arithmetic mean is calculated, it will approach

the mean or the expectation of the random variable X as the number of observations becomes very large.

An important result concerning the sum of the random variables is given by

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

That is the expectation of the sum of a group of random variables is equal to the sum of the expectations of the random variables. This result holds even if the random variables are not independent.

Variance

The arithmetic mean indicates the central tendency and provides a value around which the random variable X is distributed. Two or more random variables may have the same mean value but the deviations from this value may have different likelihoods. The information regarding the scatter of the values around the mean is provided by the variance of X, designated $V(X)$

$$V(X) = E[(X - E(X))^2] \tag{2.12}$$

$$= \sum_i (x_i - m)^2 p(x_i) \quad \text{for the discrete case}$$

and

$$= \int_{-\infty}^{+\infty} (x - m)^2 f(x)\,dx \quad \text{for the continuous case}$$

The variance is a weighted average of the values $(X\text{-}m)^2$ and therefore if large it means that the distribution of X is such that large deviations occur with a comparatively high probability. On the other hand if the values are near m with large probability, the variance is comparatively small. The variance, therefore, provides us with a measure of the spread of the distribution. Equation (2.12) can be put into another form more suitable for computation

$$V(X) = E[(X - E(X))^2]$$
$$= E(X^2) - (E(X))^2 \tag{2.13}$$

The variance is often denoted by σ^2 and has the dimensions of X^2. The square root of the variance σ has the dimensions of X and is called the standard deviation.

Covariance

Consider the two random variables X and Y. The quantities $(X - E(X))$ and $(Y - E(Y))$ are the deviations of the two random variables from their respective means. The expected value of the product of these deviations is called the covariance and is given by

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$
$$= E(XY) - E(X)E(Y) \tag{2.14}$$

In terms of the joint probability density function $f(x,y)$ of X and Y, it is given as

$$\text{Cov}(X, Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - m_x)(y - m_y) f(x, y)\,dx\,dy \tag{2.15}$$

where $f(x, y)$ is so defined that

$$P(x < X \leqslant x + dx, y < Y \leqslant y + dy) = f(x,y)dxdy$$

The covariance gives a measure of the tendency of the two random variables to vary together. If both the deviations have the same sign, then the sign of the product is positive and otherwise it is negative. Therefore, if the two variables tend to vary in harmony, the sign of the product will be positive with a larger probability and the covariance will, therefore, be positive. On the other hand if the two variables tend to vary in opposition, the sign of the covariance is negative. When the two random variables are independent, their covariance is zero. This tendency to vary together or in opposition is often measured by the dimensionless quantity called the correlation coefficient

$$C_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)\,V(Y)}} \tag{2.16}$$

which can be shown to lie in the range [-1, 1]. The correlation coefficient is the covariance normalized by the product of the standard deviations and its sign is the same as that for covariance. The statements made about the sign of the covariance, therefore, apply to the correlation coefficient too. It should be noted that correlation and non-correlation are similar to independence and interdependence but the two are not identical. If the two random variables are independent then $C_{XY} = 0$ but the reverse is not always true, i.e. independence does not necessarily follow from non-correlation. Correlation

implies dependence but the reverse is not necessarily true i.e. $C_{XY} \neq 0$ does not follow from dependence.

Moments

The expectation of a real valued function $g(X)$ of X is given by

$$\text{(i)} \quad E[g(X)] = \sum_i g(x_i)p(x_i) \qquad \text{if } X \text{ is discrete}$$

$$\text{(ii)} \quad E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)\,dx \qquad \text{if } X \text{ is continuous.} \tag{2.17}$$

One simple function whose expectation is of interest is the kth power of X, i.e. $g(x) = X^k$. The expectation of this function is called the kth initial moment of X or the distribution of X. Therefore the kth initial moment of X is given by

$$\begin{aligned} m_k(X) &= m_k \\ &= \sum_i x_i^k p(x_i) \qquad \text{for the discrete case} \\ &= \int_{-\infty}^{+\infty} x^k f(x)\,dx \qquad \text{for the continuous case.} \end{aligned} \tag{2.18}$$

A somewhat more useful concept, the kth central moment, is similarly defined by

$$\begin{aligned} M_k(X) &= M_k \\ &= \sum_i (x_i - m)^k p(x_i) \qquad \text{if } X \text{ is discrete} \\ &= \int_{-\infty}^{+\infty} (x-m)^k f(x)\,dx \qquad \text{if } X \text{ is continuous.} \end{aligned} \tag{2.19}$$

It can be seen that the mean is the first initial moment, the mean square value is the second initial moment and the variance is the second central moment.

In a similar manner, the mixed initial moment for the multi-dimensional distribution can be defined as

$$m_{k_1, k_2, \ldots, k_n}(X_1, X_2, \ldots, X_n) = E(X_1^{k_1} \quad X_2^{k_2} \ldots X_n^{k_n}) \tag{2.20}$$

and the mixed central moment is given by

$$\begin{aligned} &M_{k_1, k_2, \ldots, k_n}(X_1, X_2, \ldots, X_n) \\ &\quad = E[(X_1 - E(X_1))^{k_1}(X_2 - E(X_2))^{k_2} \ldots (X_n - E(X_n))^{k_n}] \end{aligned} \tag{2.21}$$

The covariance is therefore the first central mixed moment of X and Y.

Coefficients of Skewness and Excess

It can be seen from Equation (2.19) that odd moments vanish for a distribution which is symmetrical about the mean value. The third moment can, therefore, be used as a measure of the asymmetry of the distribution. Asymmetry is more conveniently measured as a dimensionless quantity and the asymmetry coefficient or skew coefficient is given by

$$A = M_3/(M_2)^{3/2} \tag{2.22}$$

The coefficient of excess also gives information about the form of distribution by comparing it with the normal probability density function near its mode and is defined by

$$G = M_4/(M_2)^2 - 3 \tag{2.23}$$

$M_4 = 3M_2^2$ for the normal distribution and therefore the coefficient of excess is zero. In the case of a distribution having the same variance as the normal, $G > 0$ indicates that the distribution has a sharper peak than the normal distribution and similarly $G < 0$ indicates a comparatively flatter peak.

Transform Methods

Operational or transform methods are employed for transforming the problem into a functional form which at first glance appears to have nothing to do with the original problem but which often facilitates its solution. They are used in many branches of mathematics especially in solving differential equations. One technique of importance in probability theory is the method of characteristic functions. The characteristic function of a random variable X is defined by

$$\begin{aligned} \phi_X(\theta) &= E[e^{i\theta X}] \\ &= \int_{-\infty}^{\infty} \exp(i\theta x) f(x)\,dx \end{aligned} \tag{2.24}$$

where

$$i = \sqrt{-1}$$

and $f(x)$ = the probability density function of X

The probability density function can be found from the characteristic function by the inversion formula

$$f(x) = \int_{-\infty}^{+\infty} \phi(\theta) \exp(-i\theta x)\, d\theta \tag{2.25}$$

There is a one to one correspondence between the characteristic function and the probability density function. Therefore, two characteristic functions which are equal at all points are the characteristic functions of the same distribution function and vice versa.

Differentiating (2.24) k times

$$\phi^{(k)}(\theta) = \frac{d^k \phi(\theta)}{d\theta^k} = i^k \int_{-\infty}^{\infty} x^k \exp(i\theta x) f(x)\, dx$$

This is with the assumption that the kth derivative exists. As $\theta \to 0$

$$\begin{aligned}\phi^{(k)}(0) &= i^k \int_{-\infty}^{+\infty} x^k f(x)\, dx \\ &= i^k m_k\end{aligned}$$

Therefore the kth moment of the random variable X, if it exists, can be obtained from the kth derivative of the characteristic function evaluated at zero

$$m_k = i^{-k} \phi^{(k)}(0) \tag{2.26}$$

An important result in probability theory is that the characteristic function of the sum of several independent random variables is the product of the characteristic functions of the random variables. Let

$$Y = X_1 + X_2 + \ldots + X_n$$

where X_i are the independent random variables. Then

$$\phi_Y(\theta) = E(e^{i\theta Y}) = E(e^{i\theta(X_1 + X_2 + \ldots + X_n)}).$$

Since the random variables are independent

$$\begin{aligned}\phi_Y(\theta) &= E(e^{i\theta X_1}) E(e^{i\theta X_2}) \ldots E(e^{i\theta X_n}) \\ &= \phi_{X_1}(\theta) \phi_{X_2}(\theta) \ldots \phi_{X_n}(\theta)\end{aligned} \tag{2.27}$$

Another transform often used is the moment generating function, defined for all real numbers by

$$\begin{aligned}\psi(\theta) &= E(e^{\theta X}) \\ &= \int_{-\infty}^{+\infty} e^{\theta x} f(x)\, dx \qquad \text{for the continuous case.} \\ &= \sum_i e^{\theta x_i} p(x_i) \qquad \text{for the discrete case.}\end{aligned} \tag{2.28}$$

As in the case of the characteristic function, the following results are important

$$m_k = \psi^{(k)}(0) \tag{2.29}$$

and

$$\psi_Y(\theta) = \psi_{X_1}(\theta) \psi_{X_2}(\theta) \ldots \psi_{X_n}(\theta) \tag{2.30}$$

when

$$Y = X_1 + X_2 + \ldots + X_n$$

X_j being independent random variables.

The Expression (2.29) is suitable for calculating the kth initial moment and the central moments can be calculated from the initial moments. The moment generating function about the mean m or in fact any other point can also be defined as

$$\begin{aligned}\psi_m(\theta) &= E(e^{(X-m)\theta}) \\ &= e^{-m\theta} \psi(\theta)\end{aligned} \tag{2.31}$$

The kth moment about the mean can be calculated from its kth derivative evaluated at zero, i.e.

$$M_k = \psi_m^{(k)}(0) \tag{2.32}$$

The main advantage of the characteristic function over the moment generating function is that it always exists whereas the moment generating function may not exist.

For an integer valued discrete random variable X having a probability mass function p_i, the moment generating function can be simplified by substituting $z = e^{\theta}$

$$\psi(Z) = \sum_i z^i p_i \tag{2.33}$$

This function is called the probability generating function of X or the Z transform of X. It can be shown that

$$p_k = \frac{\psi^{(k)}(0)}{k!} \tag{2.34}$$

$$m_k = \psi^{(k)}(1) \tag{2.35}$$

Reliability modelling is usually concerned with non-negative random variables and for these the Laplace transform is very useful. In general if $g(t)$ is a real function of a variable t defined for $t > 0$, then the Laplace transform of this function is given by

$$L[g(t)] = \bar{g}(s) = \int_0^{+\infty} g(t)\, e^{-st}\, dt \tag{2.36}$$

where s is a complex variable. The inverse Laplace transform is defined by

$$L^{-1}[\bar{g}(s)] = g(t) = \frac{1}{2\pi} \int_{c-i\infty}^{c+i\infty} \bar{g}(s)\, e^{st}\, ds$$

when $i = \sqrt{-1}$

In practice it is seldom necessary to perform this contour integration and the inverse is calculated by expanding $\bar{g}(s)$ into partial fractions and using the tables of Laplace transforms. The following results are of importance with respect to this book.

1. The Laplace transform of the derivative of a function $g(t)$ whose Laplace transform is $\bar{g}(s)$ is given by

$$L\left[\frac{dg(t)}{dt}\right] = s\bar{g}(s) - g(0^+) \tag{2.37}$$

Here $g(0^+)$ is the limit of $g(t)$ as t approaches zero from positive values. Very often this will be written simply as $g(0)$.

2. The Laplace transform of an integral is given by

$$L\left[\int_0^t g(u)\, du\right] = \frac{\bar{g}(s)}{s} \tag{2.38}$$

3. The Initial Value Theorem states

$$g(0^+) = \lim_{t \to 0} g(t) = \lim_{s \to \infty} s\bar{g}(s) \tag{2.39}$$

4. The Final Theorem states that

$$g(\infty) = \lim_{t \to \infty} g(t) = \lim_{s \to 0} s\bar{g}(s) \tag{2.40}$$

if the limit $\lim_{t \to \infty} g(t)$ exists

While considering the Laplace transform of a probability density function, it should be observed that except for the sign of s it is the same as the moment generating function. Therefore

$$L[f(x)] = \bar{f}(s) = E(e^{-sX})$$

As $f(x)$ is a probability density function it follows from (2.36) that

$$\bar{f}(0) = 1$$

Using the result (2.38), the Laplace transform of the probability distribution function is given by

$$\bar{F}(s) = \frac{\bar{f}(s)}{s} \tag{2.41}$$

and the Laplace transform of the survivor function is given by

$$\bar{\mathcal{F}}(s) = \frac{1 - \bar{f}(s)}{s} \tag{2.42}$$

The $\bar{f}(s)$ can be written as

$$\bar{f}(s) = E\left[\sum_{k=0}^{\infty} (-1)^k \frac{s^k X^k}{k!}\right]$$

$$= \sum_{k=0}^{\infty} (-1)^k \frac{s^k}{k!} m_k \tag{2.43}$$

The kth initial moment of X is given by the coefficient of $\frac{(-s)^k}{k!}$ in the Taylor expansion of $\bar{f}(s)$. Similarly for the kth central moment M_k, $e^{sm}\bar{f}(s)$ should be expanded.

Some Special Distributions

This book is mainly concerned with distributions of continuous random variables such as the time to failure and time to repair. Any function which is non-negative and whose integral over its range equals one, can be a probability density function. There are, however, some special mathematical forms which are often used in reliability modelling as probability density functions. The following section examines these special distributions.

1 *Exponential Distribution*

A non-negative continuous random variable is said to have a negative exponential distribution if it has a probability density function defined as

$$f(x) = \rho\, e^{-\rho x} \tag{2.44}$$

where ρ is some positive constant. The corresponding distribution function is given by

$$F(x) = \int_0^x f(u)\,du = 1 - e^{-\rho x} \tag{2.45}$$

The survivor function is

$$R(x) = \int_x^{\infty} f(u)\,du = e^{-\rho x} \tag{2.46}$$

The hazard function is

$$\phi(x) = \frac{f(x)}{R(x)} = \rho \tag{2.47}$$

The Laplace transform of this distribution is given by

$$\bar{f}(s) = L[\rho\, e^{-\rho x}] = \frac{\rho}{\rho + s}$$

$$\text{The mean} = (-1)\bar{f}^{(1)}(s)|_{s=0} = \rho(\rho + s)^{-2}|_{s=0} = \frac{1}{\rho} \tag{2.48}$$

It can be seen from Expressions (2.47) and (2.48) that the hazard rate for the exponential distribution is constant and equals the reciprocal of the mean of the random variable having this distribution. Since the hazard rate uniquely determines the probability density function, it follows that if the hazard is constant then

$$f(x) = \rho \exp\left[-\int_0^x \rho\,du\right] = \rho\, e^{-\rho x}$$

That is the distribution is exponential. It is worth repeating that only the exponential distribution has the property that the hazard rate is constant and is equal to the reciprocal of the mean. The second initial moment is

$$m_2 = (-1)^2 \bar{f}^{(2)}(s)|_{s=0} = \frac{2}{\rho^2}$$

and therefore the variance

$$M_2 = \frac{2}{\rho^2} - \frac{1}{\rho^2} = \frac{1}{\rho^2} \tag{2.49}$$

The standard deviation is $\frac{1}{\rho}$ and the coefficient of variation equals one. The graphs of the probability density, probability distribution and the hazard function are shown in Fig. 2.1. The exponential law corresponds to the maximum randomness of the lifetimes of the components. The life of a component observing the exponential law has the interesting property that the previous operating time of the component does not effect the residual or remaining lifetime distribution. Consider that a component whose lifetime is represented by an exponentially distributed random variable X is operated up to

time t then the distribution function of $Y = X - t$ is given by

$$\begin{aligned} F_Y(x) &= P[X - t \leqslant x \mid X > t] \\ &= P[X \leqslant x + t \mid X > t] \\ &= \frac{P[t < X \leqslant x + t]}{P[X > t]} \\ &= \frac{\int_t^{x+t} \rho e^{-\rho u}\, du}{e^{-\rho t}} \\ &= 1 - e^{-\rho x} \\ &= F_X(x) \end{aligned}$$

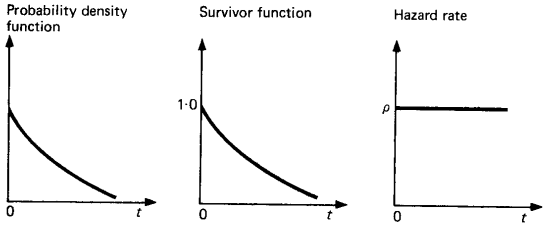


Fig. 2.1 Characteristics of the exponential distribution

It can, therefore, be seen that the distribution of the residual life time of the component is independent of the time for which the component has been operating. It is as if the component forgets how long it has been operating and the breakdown occurs not because of gradual deterioration but a randomly occurring failure. The negative exponential is the only probability distribution having this memory loss property. Its proof is left to the reader in Exercise 1.

The exponential distribution bears an important relationship with the Poisson distribution that if the number of events per unit time is given by a Poisson distribution then the distribution of the interevent time is exponential. Suppose that a component on failure is replaced by an identical component and the number of failures per unit time is λ, then if failure occurs according to the Poisson law, the distribution of failures in time t is given by

$$p_k(t) = P[\text{Number of failures} = k] = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

That no failure occurs during the interval $(0, t)$ is equivalent to having the inter failure time greater than t, i.e.

$$1 - F(t) = p_0(t) = e^{-\lambda t}$$

Differentiating both sides with respect to t

$$f(t) = \lambda e^{-\lambda t}$$

It can therefore be seen that for Poisson failures, the distribution of the time between failures is exponential.

The exponential distribution has desirable mathematical properties and is used widely for representing operating and repair times. The data collected in connection with complex equipment shows that their time of operation is indeed well described by the exponential law. It can be shown mathematically that if a complex piece of equipment has a large number of stochastically independent components such that each component is replaced immediately on failure and every component causes equipment failure, than after a long time the distribution of operating time of the equipment is well approximated by the exponential law. Though the exponential law generally holds for the operating time during the useful life of the equipment, no single hypothesis seems to exist for the repair time. It is generally realized that an exponential representation for the repair time may not be valid in many cases. When the operating time is exponentially distributed and its mean value is much larger than the mean down time, then the steady state values are not generally significantly affected by the nature of the repair time distribution. The question of non-exponential distributions for down times is examined in detail in Chapter 6.

2 *Normal Distribution*

A continuous random variable is said to be normally distributed if its probability density function is of the form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}\, dx, \quad -\infty < x < \infty \tag{2.50}$$

where σ is positive and m is any constant. The graph of (2.50) is bell-shaped and is symmetrical about $x=m$ as shown in Fig. 2.2. In the special case when $m = 0$ and $\sigma=1$, this function is called the standard normal density function.

The probability distribution for the normally distributed random variable is given by

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(u-m)^2/2\sigma^2} du \tag{2.60}$$

This function cannot be expressed in closed form in terms of familiar elementary functions. The function $F(x)$ can be transformed into the standard normal distribution by a simple change of variable

$$z = \frac{u-m}{\sigma}$$

Expression (2.60) becomes

$$F(x) = \frac{1}{2\pi} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-z^2/2} dz$$

The values of the standard normal probability distribution function are obtained by numerical integration and are tabulated in every elementary statistics book. The use of these tables is left as an exercise to the reader. The moment generating function of a normally distributed random variable is

$$\psi(\theta) = E(e^{\theta X})$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{x\theta} e^{-(x-m)^2/2\sigma^2} dx$$

Making the substitution

$$z = \frac{x-m}{\sigma} \text{ so that } x = \sigma z + m \text{ and } dx = \sigma dz$$

$$\psi(\theta) = e^{m\theta} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(z^2 - 2\sigma\theta z)/2} dz$$

$$= e^{m\theta + \sigma^2\theta^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(z-\sigma\theta)^2/2} dz$$

Making the substitution $w = z - \sigma\theta$, $dz = dw$

$$\psi(\theta) = e^{m\theta + \sigma^2\theta^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-w^2/2} dw$$

$$= e^{m\theta + \sigma^2\theta^2/2} \tag{2.61}$$

$$E(X) = \text{mean} = \psi^{(1)}(\theta)|_{\theta=0} = (m + \sigma^2\theta)\, e^{m\theta + \sigma^2\theta^2/2}|_{\theta=0}$$

$$= m$$

$$E(X^2) = \psi^{(2)}(\theta)|_{\theta=0} = [(m + \sigma^2\theta)^2 + \sigma^2]\, e^{m\theta + \sigma^2\theta^2/2}|_{\theta=0}$$

$$= m^2 + \sigma^2$$

Therefore

$$\text{Var}(X) = E(X^2) - (E(x))^2 = \sigma^2$$

The parameters m and σ, therefore, represent the mean and standard deviation of the normal distribution. The domain of the normal distribution, extends from $-\infty$ to $+\infty$. The operating times of the components are limited, however, to the positive values. If a normal distribution is to be used for modelling the operating or down times of the components, then its truncated version can be used. The truncated normal distribution is given by

$$f(x) = \frac{1}{a\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}, \quad x \geqslant 0 \tag{2.62}$$

where a is the normalizing factor and is

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{m}{\sigma}}^{\infty} e^{-u^2/2}\, du$$

The survivor function is given by

$$R(x) = \int_{x}^{\infty} f(u)\, du$$

and the hazard function

$$\phi(x) = \frac{f(x)}{R(x)}$$

The hazard rate of the normal distribution is a monotonically increasing function and the graphs of the probability density function, probability

distribution function and the hazard rate are shown in Fig. 2.2

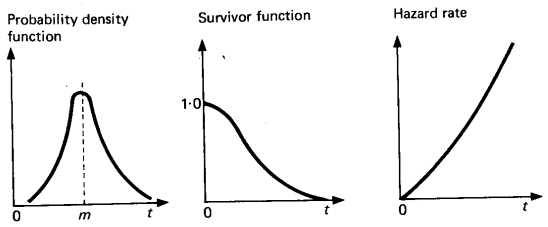


Fig. 2.2 Characteristics of the normal distribution

3 *Log-normal Distribution*

A non-negative random variable X is said to have a log-normal distribution with parameters m and σ if $Y = \log X$ is normally distributed with parameters m and σ. Since Y is a non-decreasing function of X

$$P(X \leqslant x) = P(Y \leqslant y)$$

or

$$F_X(x) = F_Y(y)$$

Differentiating both sides

$$f_X(x) = f_Y(y)\frac{dy}{dx}$$

Therefore if

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-m)^2/2\sigma^2}$$

Then

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\log x - m)^2/2\sigma^2}, \quad x \geqslant 0 \tag{2.63}$$

which is the probability density function of X having a log-normal distribution. The corresponding cumulative distribution function is

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^x e^{-(\log u - m)^2/2\sigma^2} \frac{du}{u}$$

Making the substitution

$$z = \frac{\log u - m}{\sigma}$$

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\frac{\log x - m}{\sigma}} e^{-z^2/2}\, dz$$

which is the standard normal probability integral. The hazard rate is given by

$$\phi(x) = \frac{f(x)}{1 - F(x)}$$

The hazard rate of a log-normal distribution can be shown to increase to a maximum value and then decrease to zero as $x \to 0$. The log-normal distribution does not seem to be physically suited to model component lifetimes. It does however, seem to provide a reasonable fit for many component repair times.

The expression for the kth initial moment is easily derived as

$$m_k(X) = E(X^k) = E(e^{kY}) = e^{mk + \sigma^2 k^2/2} \tag{2.64}$$

The mean is

$$m_X = e^{m + \sigma^2/2} \tag{2.65}$$

and the variance

$$\sigma_X^2 = e^{2m + 2\sigma^2} - e^{2m + \sigma^2} \tag{2.66}$$

The graphs of the probability density function, probability distribution function and hazard rate are given in Fig. 2.3

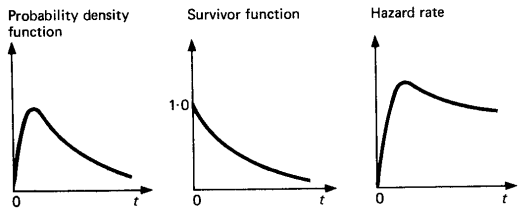


Fig. 2.3 Characteristics of the log-normal distribution

4 *Weibull Distribution*

The Weibull distribution is defined by the following probability density function

$$f(x) = \alpha\rho(\rho x)^{\alpha-1}\,e^{-(\rho x)^{\alpha}}, \quad x \geqslant 0 \tag{2.67}$$

The survivor function

$$R(x) = \int_x^{\infty} \alpha\rho(\rho u)^{\alpha-1}\,e^{-(\rho u)^{\alpha}}du$$

$$= e^{-(\rho x)^{\alpha}} \tag{2.68}$$

and therefore the hazard function

$$\phi(x) = \alpha\rho(\rho x)^{\alpha-1} \tag{2.69}$$

The hazard rate increases with x for $\alpha > 1$ and decreases for $\alpha < 1$. The graphs of the probability density, probability distribution and the hazard rate of a Weibull distribution are given in Fig. 2.4. The kth moment of a Weibull distribution can be calculated by

$$E(X^k) = \int_0^{\infty} x^k \alpha\rho(\rho x)^{\alpha-1}\,e^{-(\rho x)^{\alpha}}dx$$

Substituting $z = (\rho x)^{\alpha}$

$$m_k = \frac{1}{\rho^k}\int_0^{\infty} z^{k/\alpha}\,e^{-z}\,dz$$

$$= \frac{1}{\rho^k}\Gamma\left(\frac{k}{\alpha}+1\right) \tag{2.70}$$

where $\Gamma(a)$ is a gamma function defined by

$$\Gamma(a) = \int_0^{\infty} z^{a-1}e^{-z}\,dz$$

When a is a positive integer

$$\Gamma(a) = (a-1)!$$

The mean value

$$m = \frac{\Gamma(1+1/\alpha)}{\rho} \tag{2.71}$$

The second initial moment

$$= \frac{\Gamma(1+2/\alpha)}{\rho^2}$$

and therefore variance

$$= \frac{\Gamma(1+2/\alpha) - (\Gamma(1+1/\alpha))^2}{\rho^2} \tag{2.72}$$

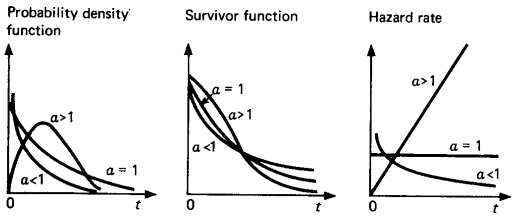


Fig. 2.4 Characteristics of the Weibull distribution

The Weibull distribution is often used in connection with mechanical components and has been used to model fatigue failure and the failure of ball bearings. It has also been recently used to describe the repair time of electric power generating units.

5 *Gamma distribution*

A non-negative continuous random variable X is said to be gamma distributed if its probability density function is given by

$$f(x) = \frac{\rho(\rho x)^{\alpha-1}\,e^{-\rho x}}{\Gamma(\alpha)} \tag{2.73}$$

This density is a function of parameters ρ and α both of which are positive constants. When α is an integer equal to a, then this distribution is also called the special Erlangian distribution,

$$f(x) = \frac{\rho(\rho x)^{a-1}\,e^{-\rho x}}{(a-1)!}$$

The probability distribution function is given by

$$F(x) = \int_0^x \frac{\rho(\rho u)^{\alpha-1} e^{-\rho u}}{\Gamma(\alpha)} du$$

Putting $\rho u = z, du = \frac{1}{\rho} dz$

$$F(x) = \frac{1}{\Gamma(\alpha)} \int_0^{\rho x} z^{\alpha-1} e^{-z} dz$$

The function $\int_0^{\rho x} z^{\alpha-1} e^{-z} dz$ is called the incomplete gamma function. When α is an integer a (special Erlangian distribution), it can be shown by integrating by parts that

$$F(x) = 1 - \sum_{k=0}^{a-1} \frac{(\rho x)^k e^{-\rho x}}{k!}$$

The Laplace transform of a gamma function is given by

$$\bar{f}(s) = \left(\frac{\rho}{\rho + s}\right)^{\alpha}$$

Therefore the mean $= (-1)\bar{f}^{(1)}(s)|_{s=0}$

$$= \frac{\alpha}{\rho}$$

The second initial moment $= (-1)^2 \bar{f}^{(2)}(s)|_{s=0}$

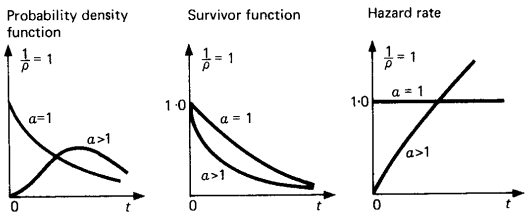


Fig. 2.5 Characteristics of the gamma distribution

$$= \frac{\alpha^2}{\rho^2} + \frac{\alpha}{\rho^2}$$

and the variance

$$= E(X^2) - (E(X))^2$$

$$= \frac{\alpha}{\rho^2}$$

Consequently the standard deviation is $\sqrt{\alpha}/\rho$ and the coefficient of variation is $1/\sqrt{\alpha}$. Some typical cases of the gamma function are shown in Fig. 2.5. The hazard rate is non-decreasing for $\alpha > 1$ and is bounded by ρ. For $0 < \alpha < 1$, the hazard rate is non-increasing and approaches zero as $x \to \infty$.

The gamma distribution is very useful due to its simplicity and also because it can be used to approximate many empirical distributions. A more detailed treatment of its usefullness is given in Chapter 6 while discussing the device of stages. A special case is when in Equation (2.73) $\rho = \frac{1}{2}$ and $\alpha = \frac{n}{2}$ where n is an integer. This distribution is the χ^2 (chi squared) distribution with n degrees of freedom.

Stochastic Processes

Consider the example of two transmission links again. $Z(t)$ defines the state of this system at time t. There is a random variable associated with each value of t. The family of random variables $(Z(t), t \geqslant 0)$ is called a stochastic process. The values assumed by the process are called the states of the system and the set of all possible states is called the state space. The set of the possible values of the indexing parameter is called the parameter space. The indexing parameter in the above example is time but other kinds of indexing parameter such as space are also possible. For example the number of fibres at a point on a yarn can be considered a stochastic process with the length of the yarn as the parameter. This book is, however, generally concerned with time dependent processes and therefore time is the basic indexing parameter. The stochastic process is, therefore, considered as a model of a system which develops in time according to probabilistic laws.

Consider two identical and independent systems of two transmission links. The state space for each system is defined as follows:

$Z(t) = 0$	both links up
$Z(t) = 1$	one link up
$Z(t) = 2$	no link up

Assume that at time zero both the links are up, and the systems are observed

for twelve hours each. The state of each system as a function of time is shown in Fig. 2.6. These two observations are called the two independent realizations of the stochastic process. Realizations or sample paths could be obtained either by observing identical and independent systems or they could be constructed by appropriately using a table of random numbers or some equivalent randomizing device. The realizations could differ from one another in detail and in fact these differences are typical of random phenomenon. The idea of realizations often helps to provide a better insight in certain reliability problems.

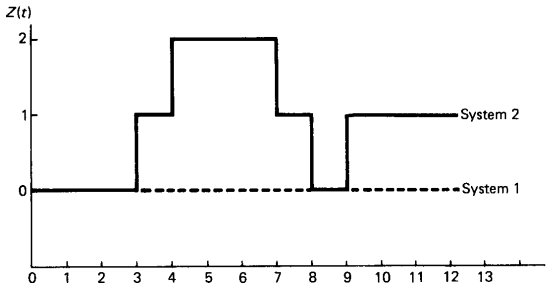


Fig. 2.6 Realizations of a stochastic process

The stochastic processes may be classified on the basis of the nature of the state space and the parameter space. The following are the four possible combinations:

1 *Discrete state and parameter spaces*
An example of this nature would be the number of successful missile flights in a missile firing scheme. The indexing parameter would be the number of missiles fired.

2 *Discrete state space and continuous parameter space*
Most of this book is concerned with stochastic processes of this kind. The states of a system of components with time as the indexing parameter is an example of this type of stochastic process.

3 *Continuous state space and discrete parameter space*
An example of this is the load on the electric power system observed every hour. The solution as such requires more sophisticated tools. These problems can, however, be often idealized by Category 1.

4 *Continuous state space and continuous parameter space*
An example of this nature is the storage in a dam observed as a function of time. Such problems can be idealized by Categories 1 or 2.

Probability Distributions

A stochastic process is defined for a set of points which may be either integer coordinates ($n = 0, 1, 2, \ldots$) or an interval of real time ($0 \leqslant t$ or $-\infty < t < \infty$). At a particular point, the stochastic process is a simple random variable. Assuming k arbitrary time points $t_l, t_m, t_n \ldots t_l < t_m < t_n < \ldots$, there are k random variables $Z(t_l), Z(t_m), Z(t_n) \ldots$. In the discrete time case these may be denoted by $Z_l, Z_m, Z_n, \ldots$. The stochastic process is completely determined in principle if the joint distribution of $Z_l, Z_m, Z_n \ldots$ is known for every k and every choice of l, m and n. In practice, however, it is rarely possible to work with these joint distributions and most of the information of interest can be obtained from transition distribution functions. One property of basic interest in the reliability evaluation of a system is the probability distribution of Z_n for the discrete time case and $Z(t)$ in the continuous time case.

Consider first the discrete time case. The stochastic process is said to be independent if

$$P(Z_n = x \mid Z_m = y,\ Z_l = z, \ldots) = P(Z_n = x)$$

This means that the probability distribution of Z_n is independent of the present and the past history of the process. A slight weakening of this condition leads to the well known class of stochastic processes called the Markov process. In this case

$$P(Z_n = x \mid Z_m = y,\ Z_l = z, \ldots) = P(Z_n = x \mid Z_m = y)$$

That is, the probability distribution of Z_n depends on the latest of the time points and none prior to that. For this reason, the Markov process is sometimes called memoryless. The Markov property essentially states that once the state occupied at a time point is known, the previous history of the process is not involved in determining the subsequent probability distributions. The Chapman–Kolmogorov equation gives the conditional probability density function for this process

$$P(Z_n = x \mid Z_l = z) = \int_{-\infty}^{\infty} P(Z_m = y \mid Z_l = z) P(Z_n = x \mid Z_m = y)\, dy \qquad (2.74)$$

Equation (2.74) is for the continuous state space, discrete time case. The corresponding equation for continuous state space and continuous time can be written as

$$P(Z(t_n) \leqslant x \mid Z(t_l) = z) =$$

$$\int_{-\infty}^{\infty} P(Z(t_n) \leqslant x \mid Z(t_m) = y)\, dP(Z(t_m) \leqslant y \mid Z(t_l) = z) \tag{2.75}$$

Equations (2.74) and (2.75) in their general form are rarely used in practice. They, however, convey the fundamental idea of recursively building the conditional probability density function over the long time interval (l, n) from those over the shorter time intervals (l, m) and (m, n). If the conditional probability density function depends only on the distance $t_m - t_l$ and not on t_n and t_l the stochastic process is called time homogenous.

The remainder of this chapter discusses the 'discrete state space and discrete time' and 'discrete state space and continuous time' Markov processes. Most of the reliability modelling falls into the latter case. It is, however, sometimes convenient to idealize the continuous time by discrete time processes. The next chapter discusses the stochastic processes from the point of view of the frequency balance.

Markov Chains

This section considers a Markov process with discrete state space and discrete parameter space. Equation (2.74) can be simplified when the state space is discrete

$$P(Z_n = x \mid Z_l = z) = \sum_y P(Z_n = x \mid Z_m = y) P(Z_m = y \mid Z_l = z)$$

where x, y, z now denote the discrete states of the system. This equation develops the conditional probability density function over the longer interval from those of shorter interval. In practice, however, it is usual to work with one step transition probabilities. In this case the Markov property states

$$P(Z_n = x \mid Z_{n-1} = y, Z_{n-2} = z, \ldots) = P(Z_n = x \mid Z_{n-1} = y) \tag{2.76}$$

If this one step transition probability is independent of n, i.e.

$$P(Z_n = x \mid Z_{n-1} = y) = P(Z_m = x \mid Z_{m-1} = y)$$

the process is time homogenous and transition probabilities are termed

stationary. The one step stationary probabilities will be denoted by p_{ij}, which is the probability of transiting from state i to state j in one step. It is easy to see that

$$\sum_j p_{ij} = 1$$

The n step transition probability $p_{ij}^{(n)}$ can be similarly defined as

$$p_{ij}^{(n)} = P(Z_{m+n} = j \mid Z_m = i)$$

The one step transition probabilities can be arranged in matrix form

$$P = (p_{ij})$$

This matrix is called the transition matrix and its ijth entry is the probability of transiting from state i to state j in one step. Each row sums to unity. A matrix which has non-negative entries with each row summing to 1 is called a stochastic matrix. Equation (2.75) can be written in terms of the single step transition probabilities as follows

$$p_{ij}^{(2)} = \sum_k p_{ik} p_{kj} \tag{2.77}$$

Arranged in matrix form

$$P^{(2)} = P.P = P^2 \tag{2.78}$$

The matrix of two step transition probabilities can be found by squaring the transition matrix. It can be easily seen that

$$P^{(n)} = P^n \tag{2.79}$$

In practice, interest is usually focussed on the probability distribution of Z_n given the initial state of the system. The initial state of the system is defined by an initial probability vector

$$p^{(0)} = (p_0, p_1, p_2, \ldots)$$

The vector of state probabilities after n steps is found by

$$p^{(n)} = p^{(0)} P^n \tag{2.80}$$

Written in the component form, this formula becomes

$$p_j^{(n)} = \sum_k p_k p_{kj}^{(n)} \tag{2.81}$$

where

$p_j^{(n)}$ = The probability of being in state j after n steps.

p_k = The probability of being in state k at the start.

and $p_{kj}^{(n)}$ = The probability of being in state j in n steps starting in state k.

Example: A person is practising firing. If he misses, he becomes nervous and the probability of the next shot being a hit reduces to ½, but a hit bolsters his confidence and the chance of the next shot being a hit increases to ¾. If the initial shot is a hit, what is the probability of a hit on the fourth shot? Also calculate this probability for the initial shot being a miss.

Designating the hit by 0 and miss by 1, the object is to find the probability distribution for Z_4. The state transition diagram is shown in Fig. 2.7.

From Equation (2.80)

$$p^{(3)} = p^{(0)} P^3$$

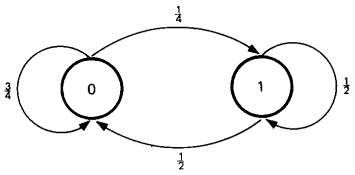


Fig. 2.7 State transition diagram

For the first shot being a hit

$$p^{(3)} = (1 \quad 0)\begin{bmatrix} \frac{43}{64} & \frac{21}{64} \\ \frac{21}{32} & \frac{11}{32} \end{bmatrix}$$

$$= \left(\frac{43}{64} \quad \frac{21}{64}\right)$$

That is the probability of a hit on the fourth shot, the first shot being a hit, is 43/64. If, however, the initial shot is a miss

$$p^{(3)} = (0 \quad 1)\begin{bmatrix} \frac{43}{64} & \frac{21}{64} \\ \frac{21}{32} & \frac{11}{32} \end{bmatrix}$$

$$= \left(\frac{21}{32} \quad \frac{11}{32}\right)$$

The probability of a hit in this case is slightly less than the previous case.

It should be noted here that the (ij)th entry of $P^{(n)}$ represents the probability of being in the jth state after n steps, given the system started in state i. The states of a discrete Markov chain can be classified into the following types.

If the states i and j can be reached from each other in a finite number of steps, they are said to communicate. The set of states in which each pair of states communicate and which once entered cannot be left is called a closed communicating class. This is also called an ergodic set of states. On the other hand, a set of states in which every state can be reached from every other state is called a transient set. The discrete chain in which every state can be reached from every other state is termed irreducible or ergodic. In other words the states form a single closed communicating set. An ergodic chain in which each state can be entered only at certain periodic intervals is called cyclic or periodic chain. If a state exhibits this characteristic, then the state is termed periodic or cyclic. A discrete Markov chain which is ergodic and a-periodic is called a regular chain. The periodic chains and states are troublesome to deal with, but fortunately reliability problems are most frequently described by regular chains.

Equilibrium Distribution

In the firing practice example

$$P^3 = \begin{bmatrix} 0{\cdot}6719 & 0{\cdot}3281 \\ 0{\cdot}6563 & 0{\cdot}3438 \end{bmatrix}$$

$$P^6 = \begin{bmatrix} 0{\cdot}6667 & 0{\cdot}3333 \\ 0{\cdot}6665 & 0{\cdot}3335 \end{bmatrix}$$

$$P^{12} = \begin{bmatrix} 0{\cdot}6666 & 0{\cdot}3334 \\ 0{\cdot}6666 & 0{\cdot}3334 \end{bmatrix}$$

It can be seen that the entries of P^n seem to be approaching a limiting value. Is this true in all cases? The following results are stated without proof.

1. In any Markov chain which is not cyclic the limit $x_j = \lim_{n\to\infty} p_j^{(n)}$ exists.
2. In any a-periodic, irreducible Markov chain the above limit does not depend on the initial probability distribution so that

$$x_j = \lim_{n\to\infty} p_j^{(n)} = \lim_{n\to\infty} p_{ij}^{(n)}$$

3. In a finite regular Markov chain, each row approaches a stationary probability vector $\alpha = (\alpha_0, \alpha_1, \ldots)$. This is called the unique stationary probability vector of the process and

$$\alpha P = \alpha \tag{2.82}$$

This relationship is very useful for determining the limiting state probabilities (also called steady state probabilities) of the process. In the firing practice example

$$(\alpha_0 \quad \alpha_1)\begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = (\alpha_0 \quad \alpha_1)$$

i.e.

$$-\tfrac{1}{4}\alpha_0 + \tfrac{1}{2}\alpha_1 = 0 \tag{2.83}$$

$$\tfrac{1}{4}\alpha_0 - \tfrac{1}{2}\alpha_1 = 0 \tag{2.84}$$

These two equations are identical, therefore an equation of the following form can be used

$$\alpha_0 + \alpha_1 = 1 \tag{2.85}$$

From Equations (2.83) and (2.85)

$$\alpha_0 = \tfrac{2}{3}$$

and

$$\alpha_1 = \tfrac{1}{3}$$

It can be seen that these values could also be obtained by multiplying P, a large number of times.

Time Specific Behaviour

It has been shown that the n-step probability distribution of the discrete Markov chain can be found from P^n where P is the transition matrix. In determining higher powers of P, the following matrix product is often useful

$$P^n = P^{n-m}P^m \tag{2.86}$$

The multiplication of large matrices is quite unwieldy using hand calculations but easily accomplished when a digital computer is used. Though this method of matrix multiplication is quite useful, the following technique can be used for very large powers of P.

If the matrix P has N distinct real eigenvalues, then it can be proved that there exists a matrix S having an inverse S^{-1} such that

$$SPS^{-1} = D \tag{2.87}$$

The matrix P is then said to be similar to the diagonal matrix D. In the diagonal matrix all but the diagonal elements are zero. The diagonal elements of D are the eigenvalues of P and can be determined from the following relationship.

$$\det(P - dI) = 0 \tag{2.88}$$

Equation (2.87) can be rearranged as

$$SP = DS \tag{2.89}$$

The row s_i of S is termed the ith left eigenvector associated with the eigenvalue d_{ii}. Similarly rearranging (2.89)

$$PS^{-1} = S^{-1}D \tag{2.90}$$

The jth column vector of S^{-1} is termed the jth right eigenvector of S. It can be seen from (2.87) that

$$SPS^{-1}SPS^{-1} = SP^2S^{-1} = D^2$$

and by induction

$$P^n = S^{-1}D^nS \tag{2.91}$$

The nth power of P can therefore be found from the nth power of the diagonal

matrix which is easy. The difficult part, however, is to determine the eigenvalues and the associated eigenvectors of P. Several numerical techniques are available for determining these elements. In many practical problems, the basic matrix multiplication technique is quite adequate. The matrix algebra approach to the firing practice problem is as follows

$$P = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Therefore

$$(P - dI) = \begin{bmatrix} \frac{3}{4} - d & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} - d \end{bmatrix}$$

The eigenvalues can now be found by equating the determinent to zero.

$$\begin{vmatrix} \frac{3}{4} - d & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} - d \end{vmatrix} = 0$$

$$(\tfrac{3}{4} - d)(\tfrac{1}{2} - d) - \tfrac{1}{8} = 0$$

i.e.

$$4d^2 - 5d + 1 = 0$$

The roots of this equation give the eigenvalues of P

$$d_0 = 1$$

$$d_1 = \tfrac{1}{4}$$

Therefore

$$D = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$$

The next step involves determining the left and right eigenvectors of P. From Equation (2.89)

$$\tfrac{3}{4}s_{00} + \tfrac{1}{2}s_{01} = s_{00}$$

i.e.

$$-\tfrac{1}{4}s_{00} + \tfrac{1}{2}s_{01} = 0 \qquad (2.93)$$

Also

$$\tfrac{1}{4}s_{00} - \tfrac{1}{2}s_{01} = 0 \qquad (2.94)$$

$$\tfrac{1}{2}s_{10} + \tfrac{1}{2}s_{11} = 0 \qquad (2.95)$$

and

$$\tfrac{1}{4}s_{10} + \tfrac{1}{4}s_{11} = 0 \qquad (2.96)$$

Equation (2.93) and (2.94) are identical, as are Equations (2.95) and (2.96). There are now two equations and four unknowns and therefore the magnitudes of the eigenvectors cannot be uniquely determined. Assuming

$$s_{00} = s_{11} = 1$$

$$s_{01} = \tfrac{1}{2}$$

$$s_{10} = -1$$

Therefore

$$S = \begin{bmatrix} 1 & \frac{1}{2} \\ -1 & 1 \end{bmatrix}$$

It should be noted that the elements of S, in general, are not determined uniquely. Each row of S is determined up to a multiplicative constant. S^{-1} can be found by inverting S

$$S^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

Using (2.91)

$$P^n = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} (1)^n & 0 \\ 0 & (\frac{1}{4})^n \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ -1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{2}{3} + \frac{1}{3}(\frac{1}{4})^n & \frac{1}{3} - \frac{1}{3}(\frac{1}{4})^n \\ \frac{2}{3} - \frac{2}{3}(\frac{1}{4})^n & \frac{1}{3} + \frac{2}{3}(\frac{1}{4})^n \end{bmatrix}$$

For $n = 3$

$$P^3 = \begin{bmatrix} \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{64} & \frac{1}{3} - \frac{1}{3} \cdot \frac{1}{64} \\ \frac{2}{3} - \frac{2}{3} \cdot \frac{1}{64} & \frac{1}{3} + \frac{2}{3} \cdot \frac{1}{64} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{43}{64} & \frac{21}{64} \\ \frac{21}{32} & \frac{11}{32} \end{bmatrix}$$

as previously found by the matrix multiplication approach. Similarly

$$\underset{n \to \infty}{P^n} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

First Passage Times

One parameter of interest in many Markov Chain problems is the time to encounter a state for the first time. This is called the first passage time. If this state is an absorbing state or has been made an absorbing state, this is called the time of absorption. In reliability engineering this concept is used to calculate the mean time to first failure, MTTFF. As noted earlier, almost all the cases of practical interest are regular chains, i.e. chains in which all the states communicate and which are not cyclic. In these cases, the mean first passage times and their variance can be obtained from the fundamental matrix Z defined as below

$$Z = [I - [P - A]]^{-1} \tag{2.97}$$

where

I is the identity matrix

P is the transition matrix

and A is the matrix each row of which is the limiting probability vector

$$\alpha = (\alpha_0, \alpha_1, \ldots .)$$

The mean first passage time matrix $\bar{T}$ is given by

$$\bar{T} = [I - Z + UZ_d]D \tag{2.98}$$

where

$\bar{T}$ is the mean first passage time matrix such that $\bar{t}_{ij}$ represents the mean time or mean number of steps to go from state i to j

U A unit matrix, i.e. with all entries 1

Z_d Matrix resulting from Z by setting off-diagonal elements equal to zero

D Diagonal matrix such that $d_{ii} = 1/\alpha_i$

The variance of the first passage times can also be explicitly determined. Denoting the first passage time from state i to j by t_{ij}, define the matrix W

$$W = (E(t_{ij}^2))$$

This matrix can be computed from the fundamental matrix by

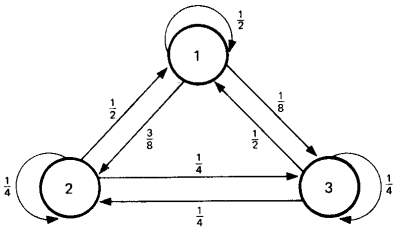
$$W = \bar{T}(2Z_dD - I) + 2[Z\bar{T} - U(Z\bar{T})_d] \tag{2.99}$$

$(\bar{ZT})_d$ is a matrix obtained by setting the off-diagonal elements of $\bar{ZT}$ equal to zero.

The variance can now be obtained using Equation (2.13)

$$V(t_{ij}) = E(t_{ij}^2) - (E(t_{ij}))^2$$
$$= w_{ij} - \bar{t}_{ij}^2$$

Example: A discrete Markov chain has the state transition diagram shown below. Find the matrix of the mean first passage times.



The transition matrix for this state space diagram is

$$P = \begin{bmatrix} \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

The vector α can be computed by

$$(\alpha_1 \quad \alpha_2 \quad \alpha_3)\begin{bmatrix} \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} = (\alpha_1 \quad \alpha_2 \quad \alpha_3)$$

On solving these equations along with $\alpha_1 + \alpha_2 + \alpha_3 = 1$

$$\alpha_1 = \tfrac{1}{2}$$
$$\alpha_2 = \tfrac{5}{16}$$
$$\alpha_3 = \tfrac{3}{16}$$

Therefore

$$A = \begin{bmatrix} \frac{1}{2} & \frac{5}{16} & \frac{3}{16} \\ \frac{1}{2} & \frac{5}{16} & \frac{3}{16} \\ \frac{1}{2} & \frac{5}{16} & \frac{3}{16} \end{bmatrix}$$

$$I-[P-A] = \begin{bmatrix} 1 & -\frac{1}{16} & \frac{1}{16} \\ 0 & \frac{17}{16} & -\frac{1}{16} \\ 0 & \frac{1}{16} & \frac{15}{16} \end{bmatrix}$$

Z can be now determined from the inverse of the above matrix

$$Z = \begin{bmatrix} 1 & \frac{1}{16} & -\frac{1}{16} \\ 0 & \frac{15}{16} & \frac{1}{16} \\ 0 & -\frac{1}{16} & \frac{17}{16} \end{bmatrix}$$

$$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{16}{5} & 0 \\ 0 & 0 & \frac{16}{3} \end{bmatrix}$$

Substituting into (2.98)

$$\bar{T} = \begin{bmatrix} 1 & \frac{14}{16} & \frac{18}{16} \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{16}{5} & 0 \\ 0 & 0 & \frac{16}{3} \end{bmatrix}$$
$$= \begin{bmatrix} 2 & \frac{14}{5} & 6 \\ 2 & \frac{16}{5} & \frac{16}{3} \\ 2 & \frac{16}{5} & \frac{16}{3} \end{bmatrix}$$

Alternative Approach to First Passage Times

The technique for calculating the mean and variance of first passage times for regular Markov chains has been illustrated. It is also possible to calculate these quantities by making state j an absorbing state and applying the theory of absorbing chains. An absorbing chain is one which once entered cannot be left. The behaviour of the stochastic process before once hitting state j will be the same as that of the original process. The first passage time from state i to state j is now the time of absorption starting from state i in the new process. The basic results for this absorbing chain can be obtained from the fundamental matrix N

$$N = [I-Q]^{-1} \tag{2.100}$$

where

N = The fundamental matrix whose n_{ik} denotes the mean number of times the process is in state k before absorption, the process having been started in state i.

Q = The matrix obtained by deleting the jth row and the jth column from matrix P of transition probabilities.

The mean first passage time from state i to j is therefore

$$\bar{t}_i = \sum_{k=1}^{N-1} n_{ik}$$

The variance column vector is given by

$$W = [2N-I]\bar{t} - \bar{t}_s \tag{2.101}$$

where

W_i = The variance of the first passage time from state i to state j (the one made into an absorbing state).

$\bar{t}$ = The column vector such that $\bar{t}_i$ is the mean first passage time from i to j.

$\bar{t}_s$ = The column vector with $\bar{t}_{si} = \bar{t}_i^2$

It can be seen that this approach gives additional information about the mean first passage time by providing the components spent in various states before once hitting state j. This method can be illustrated by application to the previous example. Determine the mean first passage times from states 1 and 2 to state 3. Truncating the third column and row

$$I - Q = \begin{bmatrix} \frac{1}{2} & -\frac{3}{8} \\ -\frac{1}{2} & \frac{3}{4} \end{bmatrix}$$

The fundamental matrix N can be found from its inverse

$$N = \begin{bmatrix} 4 & 2 \\ \frac{8}{3} & \frac{8}{3} \end{bmatrix}$$

Starting in state 1, the process visits states 1 and 2, 4 and 2 times before first hitting state 3. Therefore

$$\bar{t}_{13} = 4 + 2 = 6$$

and

$$\bar{t}_{23} = \tfrac{8}{3} + \tfrac{8}{3} = 5{\cdot}333$$

It can be seen that these entries agree with the elements of $\bar{T}$ found previously. If state 3 was considered to be the failed state of the system, then MTTFF is 6 when state 1 is taken as the initial state.

Continuous Parameter Markov Chains

Many of the problems encountered in system reliability can be modelled using continuous parameter Markov chains. The next chapter examines frequency balancing techniques as an alternative way of looking at the stochastic process.

For $u < v < t$, the Markov property for a continuous parameter Markov chain would be

$$P(Z(t) = k \mid Z(v) = j, Z(u) = i) = P(Z(t) = k \mid Z(v) = j)$$

This property is basically of the form

$$P(Z(t + x) = j \mid Z(t) = i)$$

and is termed as the probability of transition from state i to state j during the time interval t to $t+x$. If this transition probability does not depend on the initial time t but only on the elapsed time x, then the process is said to be time homogeneous. This book is primarily concerned with this class of process. The transition probability will be denoted by

$$p_{ij}(x) = P(Z(t + x) = j \mid Z(t) = i)$$

for any x. The Chapman–Kolmogorov Equation (2.75) can now be written as

$$p_{ij}(t + x) = \sum_k p_{ik}(t)\, p_{kj}(x) \tag{2.102}$$

The transition probabilities must satisfy the following conditions

$$0 \leqslant p_{ij}(x) \leqslant 1 \tag{2.103}$$

and

$$\sum_j p_{ij}(x) \leqslant 1 \tag{2.104}$$

In Equation (2.104), if $\sum_j p_{ij}(x) = 1$ for all i and x, then the process is called honest but if the inequality holds then there is non-zero probability of the process escaping to infinity and such a process is termed dishonest. This book is concerned only with honest processes. In the case of discrete parameter chains, the basic elements are the one step transition probabilities. In a continuous parameter case the equivalent elements are the limiting values, i.e. as $x \to 0$. Define the transition intensity or rate as

$$\lambda_{ij} = \left.\frac{dp_{ij}(x)}{dx}\right|_{x=0} \qquad i \neq j$$

$$= \lim_{\Delta x \to 0} \frac{p_{ij}(\Delta x) - 0}{\Delta x}$$

i.e.

$$p_{ij}(\Delta x) = \lambda_{ij}\Delta x + 0(\Delta x) \tag{2.105}$$

for $i = j$

$$\lambda_{ij} = \left.\frac{dp_{ii}(x)}{dx}\right|_{x=0}$$

$$= \lim_{\Delta x \to 0} \frac{p_{ii}(\Delta x) - 1}{\Delta x}$$

i.e.

$$p_{ii}(\Delta x) = \lambda_{ii}\Delta x + 1 + 0(\Delta x) \tag{2.106}$$

Differentiating both sides of (2.104) for equality and setting $x = 0$

$$\lambda_{ii} + \sum_{j \neq i} \lambda_{ij} = 0$$

i.e.

$$\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$$

Therefore

$$p_{ii}(\Delta x) = 1 - \sum_{j \neq i} \lambda_{ij}\Delta x + 0(\Delta x) \tag{2.107}$$

In Equation (2.105), $p_{ij}(\Delta x)$ represents the probability of transiting from state i to state j during the interval of length Δx and this is equal to $\lambda_{ij}\Delta x$ plus a term which when divided by Δx tends to zero as $\Delta x \to 0$. Equation (2.107) can be interpreted in a similar manner. Equation (2.102) can now be written for a small increment of time Δt as

$$p_{ij}(t + \Delta t) = \sum_k p_{ik}(t)p_{kj}(\Delta t)$$

$$= p_{ij}(t)p_{jj}(\Delta t) + \sum_{k \neq j} p_{ik}(t)p_{kj}(\Delta t) \tag{2.108}$$

where

$$p_{ij}(t) = P(Z(t) = j \mid Z(0) = i)$$

Substituting from (2.105) and (2.106)

$$p_{ij}(t + \Delta t) = p_{ij}(t)(1 + \lambda_{jj}\Delta t) + \sum_{k \neq j} p_{ik}(t)\lambda_{kj}\Delta t + 0(\Delta t)$$

i.e.

$$\frac{p_{ij}(t + \Delta t) - p_{ij}(t)}{\Delta t} = p_{ij}(t)\lambda_{jj} + \sum_{k \neq j} p_{ik}(t)\lambda_{kj} + \frac{0(\Delta t)}{\Delta t}$$

and as $\Delta t \to 0$

$$p'_{ij}(t) = \sum_k p_{ik}(t)\lambda_{kj}$$

If $P_i(t)$ denotes the row vector whose jth element is $p_{ij}(t)$, i.e. the probability of being in the jth state at time t given that the process was initially in state i, then the above equation can be written as

$$P'_i(t) = P_i(t)R \tag{2.109}$$

where R is the transition rate matrix whose ijth element is λ_{ij}. In a more general form Equation (2.109) becomes

$$P'(t) = P(t)R \tag{2.110}$$

where $P(t)$ has $p_{ij}(t)$ as its (ij)th element. The initial condition for (2.110) is

$$P(0) = I$$

If, however the initial state of the system is defined by a probability distribution in the form of a row vector $p(0)$, the distribution at t is given by $p(0)P$. The system of equations (2.110) is termed as the system of forward equations.

At this point it is interesting to probe a little into the significance of the transition rates. Let X_{kj} be a random variable defining the duration of state k under the condition that the next transition will be to state j. In accordance with Equation (2.7), the hazard rate is

$$\phi_{kj}(x) = \lim_{\Delta x \to 0} \frac{P[x < X_{kj} \leqslant x + \Delta x \mid x < X_{kj}]}{\Delta x}$$

i.e.

$$P[x < X_{kj} \leqslant x + \Delta x \mid x < X_{kj}] = \phi_{kj}(x)\Delta x + 0(\Delta x)$$

The left hand side can be interpreted as $p_{kj}(\Delta x)$ if the process has been in state k for time x. If the process is to be Markovian then $\phi_{kj}(x)$ must be

independent of x as the process is independent of the past. Therefore

$$p_{kj}(\Delta x) = \phi_{kj}\Delta x + 0(\Delta x)$$

Comparing with (2.105).

$$\lambda_{kj} = \phi_{kj}$$

That is the transition rate λ_{kj} is the hazard rate of the random variable defining the duration of state k under the condition of transiting to state j. The exponential is the only distribution having a constant hazard rate and therefore the random variables underlying the time homogenous Markov process must be exponentially distributed.

Although time homogenous Markov Chains are the main interest in system reliability, there is no additional difficulty in extending the above arguments to transition rates which are functions of system time, i.e. when λ_{ij} is $\lambda_{ij}(t)$. The process, however, becomes non-Markovian when the transition rates are a function of the state residence times. These processes are treated in Chapter 6. In the case when the transition rates are functions of system time t there exists a family of matrices $P(u,t)$, for $t > u$ whose elements are

$$p_{ij}(u,t) = P(X(t) = j \mid X(u) = i)$$

In this case however the transition probability depends not only on $(t-u)$ but on u as well. The system of forward equations can now be written as

$$\frac{\partial P(u,t)}{\partial t} = P(u,t)R(t) \tag{2.111}$$

This equation is called the Kolmogorov differential equation.

Transient Behaviour

Equation (2.110) is a system of linear differential equations with constant coefficients. If the eigenvalues of R are distinct, the solution of Equation (2.110) can be easily obtained in the form

$$P(t) = SD(t)S^{-1} \tag{2.112}$$

where

$D(t)$ = The diagonal matrix whose (ii)th element is $\exp\{r_i t\}$, r_i being the ith eigenvalue of R

S = The matrix formed by right eigenvectors of R

S^{-1} = The matrix formed either by inverting S or from the left eigenvalue of R

The proof of Equation (2.112) may be found in books on differential equations. In practice if t is short, the solution may be found by the following technique.

$$P'(t) = P(t)R$$

As $\Delta t \to 0^+$

$$P(t+\Delta t) = P(t) + P'(t)\Delta t$$
$$= P(t)(I + R\Delta t)$$

It time t is divided into a very large number of equal intervals Δt, so that Δt is very small ($\doteq 0$), the above expression can be written as a recursive relationship

$$P(j\Delta t) = P(\overline{j-1}\ \Delta t)[I + R\Delta t] \tag{2.113}$$

It should be noted that (2.113) implies the approximation of a Markov process in continuous time by a discrete time Markov process with steps equal to Δt. The (ij)th element of $(I + R\ \Delta t)$ is $\lambda_{ij}\ \Delta t$, i.e. the probability of transiting from state i to state j in one step of length Δt. Therefore $[I + R\Delta t]$ is a one step transition probability matrix. It can also be seen that

$$P(j\Delta t) = [I + R\Delta t]^j$$

which is the matrix multiplication technique in the discrete time case

Equilibrium Probability Distribution

As $t \to \infty$, the probability distribution of $Z(t)$ tends to an equilibrium distribution. For all processes having a finite number of mutually communicating states, the unique solution can be found by solving $(N\text{-}1)$ equations from

$$pR = 0 \tag{2.114}$$

and

$$\sum_i p_i = 1 \tag{2.115}$$

where p is a row vector whose ith element p_i is the steady state probability of being in the ith state. It can be proved that p_i equals the expected value of the proportion of a long realization spent in state i. In most cases the steady state

probabilities are the only quantities of interest.

Example: One of the processes commonly encountered in reliability studies is the two state Markov process. The state transition diagram for this process is shown below.

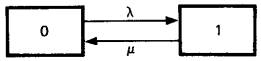


Fig. 2.8 Two-state Markov process

The transition rate matrix

$$R = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$$

Taking the Laplace of Equation (2.110)

$$sP(s) - P(0) = P(s)R$$

i.e.

$$P(s) = P(0)[sI - R]^{-1}$$

For the two state process

$$P(s) = P(0)\begin{bmatrix} s+\lambda & -\lambda \\ -\mu & s+\mu \end{bmatrix}^{-1}$$

$$= \frac{1}{s(s+\lambda+\mu)}\begin{bmatrix} s+\mu & \lambda \\ \mu & s+\lambda \end{bmatrix}$$

$$P(t) = \begin{bmatrix} \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} & \frac{\lambda}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t} \\ \frac{\mu}{\lambda+\mu} - \frac{\mu}{\lambda+\mu}e^{-(\lambda+\mu)t} & \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}e^{-(\lambda+\mu)t} \end{bmatrix}$$

The probability vector $p(t)$ of the state probabilities can be obtained by

$$(p_0(t)p_1(t)) = (p_0(0)p_1(0))P(t)$$

Therefore

$$p_0(t) = \frac{\mu}{\lambda+\mu}(p_0(0)+p_1(0)) + \left(p_0(0)\frac{\lambda}{\lambda+\mu} - p_1(0)\frac{\mu}{\lambda+\mu}\right)e^{-(\lambda+\mu)t}$$

$$= \frac{\mu}{\lambda+\mu} + \left(p_0(0) - \frac{\mu}{\lambda+\mu}\right)e^{-(\lambda+\mu)t} \quad (2.116)$$

and

$$p_1(t) = \frac{\lambda}{\lambda+\mu} + \left(p_1(0) - \frac{\lambda}{\lambda+\mu}\right)e^{-(\lambda+\mu)t} \quad (2.117)$$

As $t \to \infty$

$$\underset{t\to\infty}{p_0(t)} = p_0 = \frac{\mu}{\lambda+\mu}$$

and

$$\underset{t\to\infty}{p_1(t)} = p_1 = \frac{\lambda}{\lambda+\mu}$$

The steady state solution can also be obtained by the application of Equations (2.114) and (2.115)

$$(p_0 \quad p_1)\begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} = 0$$

Therefore

$$-\lambda p_0 + \mu p_1 = 0$$

and

$$\lambda p_0 - \mu p_1 = 0$$

One of these identical equations can be used with

$$p_0 + p_1 = 1$$

to give

$$p_0 = \frac{\mu}{\lambda+\mu}$$

and

$$p_1 = \frac{\lambda}{\lambda+\mu}$$

The probabilities p_0 and p_1 are independent of the initial condition. This seems to be intuitively true because after a long time many transitions between 0 and 1 would have taken place and therefore the effect of the initial condition tends to diminish. The probabilities p_0 and p_1 can be interpreted in two ways. The first interpretation can be in terms of an average taken over a large number of the realizations taken at a single point in time. If out of n realizations, the process is n_0 times in state 0 at a time t remote from the time origin then

$$p_0 = \lim_{n \to \infty} \frac{n_0}{n}$$

The second interpretation is in terms of the limiting proportion of time spent in state 0 in a single long realization. The parameters λ and μ are the hazard rates of exponential distributions and therefore they are the reciprocals of the mean time spent in state 0 and state 1, i.e.

$$\lambda = \frac{1}{E(X_0)}$$

$$\mu = \frac{1}{E(X_1)}$$

where X_0 and X_1 are the random variables denoting durations of 0 and 1 state.

Considering $2n$ transitions in a single long realization of the stochastic process, the process will be n times in state 0 and n times in state 1. Therefore, the proportion of time spent in state 0 in $2n$ transitions is

$$R_0^{(n)} = \frac{X_{01} + X_{02} + \ldots X_{0n}}{(X_{01} + X_{02} + \ldots + X_{0n}) + (X_{11} + X_{12} + \ldots X_{1n})}$$

Dividing both numerator and denominator by n

$$R_0^{(n)} = \frac{\bar{X}_0}{\bar{X}_0 + \bar{X}_1}$$

As the number of transitions tends to be large, the average values tend to the means (law of large numbers) and therefore

$$R_0^{(n)} = \frac{E(X_0)}{E(X_0) + E(X_1)}$$

$$n \to \infty$$

$$= \frac{\mu}{\lambda + \mu}$$

$$= p_0$$

Therefore p_0 is the limiting proportion of the time spent in state 0 in a single long realization of the two state stochastic process. A similar interpretation holds for state 1.

First Passage Times

Denote the first passage time from state i to state j by T_{ij}, i.e. this is the time to enter state j for the first time starting in state i. If the state j is now made an absorbing state, the behaviour of the new stochastic process and the original process is the same until meeting j for the first time. If $p_{ij}(t)$ is the probability of being in state j, starting in state i for the new process then

$$P(T_{ij} \leqslant t) = p_{ij}(t)$$

The probability density function $f_{ij}(t)$ can be found by differentiation

$$f_{ij}(t) = \frac{d}{dt}(P(T_{ij} \leqslant t)) = \frac{d}{dt} p_{ij}(t)$$

The Laplace transform can be obtained by

$$\bar{f}_{ij}(s) = s\bar{p}_{ij}(s) \tag{2.118}$$

The bar indicates a Laplace transform. After evaluating the right hand side, the explicit density function can be obtained by inversion. The moments of the first passage times can be found by referring to Equation (2.43).

The kth moment of the first passage time can be found by differentiating Equation (2.43), k times

$$T_{ij}^{(k)} = (-1)^k \frac{d^k}{ds^k} \bar{f}_{ij}(s)\Big|_{s=0} \tag{2.119}$$

If the absorbing state is the failed state, then the mean first passage time represents the MTTFF. The above procedure can be conveniently carried out in the matrix form. Let the states 1 to J be the elements of subset X^+ and J+1 to

N be the elements of X^-. It is required to find the first passage time to the subset X^-. The matrix of transition rates can now be partitioned as follows

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

where

R_{11} is a $J \times J$ matrix

R_{12} is a $J \times (N-J)$ matrix

R_{21} is a $(N-J) \times J$ matrix

and

R_{22} is a $(N-J) \times (N-J)$ matrix

The states $j \epsilon X^-$ are now absorbing states and therefore R_{21} and R_{22} are set to zero. Let $p(t)$ be the vector of state probabilities for an initial starting condition. This vector can be expressed as $(p_+(t)\, p_-(t))$ where $p_+(t)$ and $p_-(t)$ are the vectors containing the states $i \epsilon X^+$ and $i \epsilon X^-$ respectively. The forward differential equation now becomes

$$\frac{d}{dt}(p_+(t)p_-(t)) = (p_+(t)p_-(t)) \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix}$$

Therefore

$$p'_+(t) = p_+(t)R_{11}$$

and

$$p'_-(t) = p_+(t)R_{12}$$

Taking the Laplace transforms

$$s\bar{p}_+(s) - p_+(0) = \bar{p}_+(s)R_{11}$$

$$s\bar{p}_-(s) = \bar{p}_+(s)R_{12}$$

$p_-(0) = 0$ as the process started in $i \in X^+$. These equations can be rearranged as

$$\bar{p}_+(s) = p_+(0)[sI - R_{11}]^{-1} \tag{2.120}$$

and

$$s\bar{p}_-(s) = p_+(0)(sI - R_{11})^{-1}R_{12} \tag{2.121}$$

The probability of being in subset X^- at time t is $p_-(t)\, U_{N-k}$ where U_{N-k} is a unit vector of dimension N-k. From Equation (2.118) it can be seen that the

Laplace transform of the probability density function of the first passage time is

$$\begin{aligned} \bar{f}(s) &= s\bar{p}_-(s)U_{N-k} \\ &= p_+(0)(sI - R_{11})^{-1}R_{12}U_{N-k} \end{aligned}$$

Since the rows of the transition rate matrix sum to zero

$$R_{12}U_{N-k} = R_{11}U_k$$

Therefore

$$\bar{f}(s) = p_+(0)[sI - R_{11}]^{-1}R_{11}U_k \tag{2.122}$$

The rth initial moment can be found by Equation (2.119)

$$T^{(k)} = k!\, p_+(0)(-R_{11})^{-k}U_k \tag{2.123}$$

The mean is

$$\bar{T} = T^{(1)} = p_+(0)(-R_{11})^{-1}U_k \tag{2.124}$$

If the process started in the first state

$$\bar{T} = (1\ 0\ 0\ \ldots\ 0)(-R_{11})^{-1}U_k$$

If X^- represents the failed condition $\bar{T}$, then $\bar{T}$ is the MTTFF. It should be realized that Equation (2.124) can be derived from the theory of discrete time Markov chains by assuming that each step of the chain is $\Delta t \doteq 0$. The matrix of one step transition probabilities becomes $[I + R\,\Delta t]$ and by truncating the abosrbing states $Q = [I + R_{11}\,\Delta t]$ and therefore the fundamental matrix

$$N = [I - Q]^{-1} = \frac{1}{\Delta t}[-R_{11}]^{-1}$$

This matrix gives the number of steps spent in the different states. The time spent in the different states can be obtained by multiplying by Δt i.e. the step length. From this point on it is easy to see that

$$\bar{T} = p_+(0)(-R_{11})^{-1}U_k$$

The first passage time represents the time of entering a state or a set of states for the first time, starting in a particular state. It is sometimes necessary, however, to find the mean time spent in subset X^+ or X^-. For example, if X^+ and X^- represent the up and down states respectively, these time

parameters represent the mean up time and the mean down time. In order to calculate these quantities, it is necessary to know the probabilities of beginning X^+ in the various states, which are its elements. Denoting the steady state probabilities of being in various states of the original process by p_i, the probability of beginning X^+ in state j is

$$p_j(0) = \frac{\sum_{i \in X^-} p_i \lambda_{ij}}{\sum_{j \in X^+} \sum_{i \in X^-} p_i \lambda_{ij}}$$

In vector form

$$p_+(0) = \frac{p_- R_{21}}{p_- R_{21} U_k}$$

In the steady state

$$p_+ R_{11} + p_- R_{21} = 0$$

Therefore

$$p_+(0) = \frac{-p_+ R_{11}}{p_+ R_{11} U_k} = \frac{-p_+ R_{11}}{p_+ R_{12} U_{N-k}}$$

Substituting in (2.124), the mean stay in X^+ is

$$\begin{aligned} T^+ &= \frac{-p_+ R_{11}(-R_{11})^{-1} U_k}{p_+ R_{12} U_{N-k}} = \frac{p_+ U_k}{p_+ R_{12} U_{N-k}} \\ &= \sum_{i \in X^+} p_i \Big/ \sum_{i \in X^+} p_i \sum_{j \in X^-} \lambda_{ij} \\ &= \sum_{i \in X^+} p_i \Big/ \sum_{i \in X^-} p_i \sum_{j \in X^+} \lambda_{ij} \end{aligned} \tag{2.125}$$

In a similar manner the mean duration in state X^-

$$\begin{aligned} T^- &= \sum_{i \in X^-} p_i \Big/ \sum_{i \in X^-} p_i \sum_{j \in X^+} \lambda_{ij} \\ &= \sum_{i \in X^-} p_i \Big/ \sum_{i \in X^+} p_i \sum_{j \in X^-} \lambda_{ij} \end{aligned} \tag{2.126}$$

The mean cycle time, i.e. the time between two successive encounters of X^+ or X^-

$$\begin{aligned} T &= T^+ + T^- \\ &= 1 \Big/ \sum_{i \in X^-} p_i \sum_{j \in X^+} \lambda_{ij} \\ &= 1 \Big/ \sum_{i \in X^+} p_i \sum_{j \in X^-} \lambda_{ij} \end{aligned} \tag{2.127}$$

In the next chapter these relationships will be derived from the frequency balancing technique.

Exercises

1. X is a non-negative continuous random variable such that conditional on X being greater than a fixed value $t > 0$, the probability density function of $X - t$ is the same as the unconditional probability density function of X. Prove that X has a negative exponential probability density function.

2. Assume that X is normally distributed with $m = 0.4$ and $\sigma = 4$, find

   (a) $P(X > 1{\cdot}5)$

   (b) $P(X \leqslant 0{\cdot}5)$

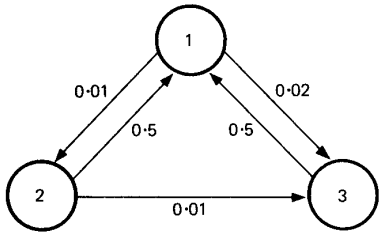
   (c) $P(-3 < X < 1)$

3. Suppose that X_i, $i = 1, 2, \ldots n$ are independent random variables, gamma distributed with parameters α_1, ρ. Prove that the random variable $\sum_{i=1}^{n} X_i$ is also gamma distributed with parameters $\sum_{i=1}^{n} \alpha_i$ and ρ. This is called the reproductive property of gamma distribution.

4. Find the 1, 2, 3 and 4 step transition probability matrix for the follow single step transition matrix. Does it exhibit any special characteristic?

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

5. The state transition diagram of a continuous time Markov chain is given below. The states 1 and 2 are working states and state 3 is failed state. Calculate
   (a) The availability, i.e. the steady state probability of being in the working state
   (b) MTTFF
   (c) Mean cycle time



References

1. U. Norayan Bhat, *Elements of Applied Stochastic Processes,* John Wiley (1972).
2. R. Billinton, *Power System Reliability Evaluation*, Gordon and Breach (1970).
3. J.A. Buzacott, Markov Approach to Finding Failure Times of Repairable Systems, *IEEE Transactions on Reliability*, **R-19**, 4, (1970).
4. A.B. Clarke and R.L. Disney, *Probability and Random Processes for Engineers and Scientists*, John Wiley (1970).
5. D.R. Cox and H.D. Miller, *The Theory of Stochastic Processes,* Methuen (1965).
6. W. Feller, *An Introduction to Probability Theory and Its Applications*, **1**. (Third Edition) John Wiley (1958).
7. J.G. Kemeny and J.L. Snell, *Finite Markov Chains,* Van Nostrand (1960).
8. E. Parzen, *Stochastic Processes*, Holden-Day (1962).
9. G.H. Sandler, *System Reliability Engineering*, Prentice-Hall (1963).