CHAPTER-II

REVIEW OF MARKOV MODELS FOR RELIABILITY ANALYSIS

2.1 INTRODUCTION

In this chapter the formulation of Markov models in the system reliability analysis is reviewed. It is in this chapter the Markovian approach and Markov models are presented and need for formulation of Markov models and various assumptions which suits to this approach in order to analyse the reliability of the system is discussed. The various Markov models, assumptions concerned in it and the reliability, availability expressions in various configuration are discussed and reviewed briefly.

The analytical and combinatorial techniques for evaluating the reliability of system is applicable both for maintained and non-maintained systems. It is assumed that the repair is done instantaneously i.e., repair time
is negligible when compared to actual operating
time of the system. In most of the real applications
it is in general not true and hence these techniques
are not adequate in such circumstances to analyse and
to study the reliability. However, a Markov approach
can overcome such a problem, and is aptly suitable in
that context. The literature on application of Markov
models are well discussed by Feller \[28\], Ryabinin \[56\]
Shooman \[58\], Sandler \[57\].

The Markov approach can be applied to the random
behaviour of the system that vary continuously or
discretely with respect to time. In the context of
study of reliability analysis, Markov models are
function of two variables viz., the state of the system
$X$ with state space $S$ (discrete) and the time of observation $t$ (continuous). Usually reliability models
would be discrete state, continuous time models
(Shooman \[58\], p-61 and Billinton et al \[16\], p-206).

In this approach the behaviour of the system is
characterised by lack of memory. That means the future states of a system are independent of all the past states except the immediately preceding one i.e., the future random behaviour of system depends on the present state of the system.

In the Markov model approach it is assumed that behaviour of the system must be same at any point of time irrespective of the instant of time being considered. Thus it is uniquely characterised as stationary or homogeneous Markov process. This means that the probability of making a transition from one state to another is same (stationary) at all times. Therefore the Markov approach is applicable to those systems, whose behaviour can be described by probability distribution that is characterised by a constant hazard rate (Billinton and Allan [16] p-206). The probability of making transition between any two states remain constant at any time point ‘t’ if the hazard rate is constant. This means that conditional probability of failure or repair during any fixed
interval time is constant. This is understood that the failure and repair characteristics of the components are associated with (negative) exponential distribution.

Under these conditions, Markov approach is well fitted to wide range of reliability problems for both repairable and non-repairable systems. The approach can be used in the case of series, parallel, stand-by configurations. The literature on Markov random processes and its applicability to the reliability analysis is discussed in Billinton [16], Cox [24], Khinchin [42], Klimov [43], Neyman [51], Pugachov [53], Riordan [55], Ryabinin [56].

2.2 MARKOV MODELLING CONCEPT

In what follows here we discuss the Markov model for the single and 2-unit systems.
2.2.1 Single Repairable Component System

In the case of single repairable component we assume that the (i) failure rate ($\lambda$) is constant i.e., if $T$ is the time to failure random variable then 'T' is assumed to follow negative exponential distribution with pdf

$$f(t, \lambda) = \lambda \exp(-\lambda t), \quad t \geq 0$$

$$\lambda > 0 \quad (2.2.1)$$

'\lambda' is the failure rate, which is the reciprocal of mean time to failure of the component.

(ii) Also repair is constant rate ($\mu$) i.e., repair time $S$, is assumed to follow negative exponential distribution with pdf

$$f(s, \mu) = \mu \exp(-\mu s), \quad s \geq 0$$

$$\mu > 0 \quad (2.2.2)$$

Thus the system is represented as a Markov process with
set of states of system.

\[ S = \{0, 1\} \]

0 = System is operative (zero component fail)
1 = System is inoperative (one component fail).

We define the states of the system at \( t=0 \) are called initial states and those representing a final or equilibrium states are called final states. The set of Markov state equations describes the probabilistic transition from the initial to the final states.

The probability of transition must obey the following rules (Shooman [58] pp.61-65)

1. The probability of transition in length of time \( dt \) from one state to another is given by \( h(t) \ dt \) where \( h(t) \) is the hazard associated with the two states.

2. The probabilities of more than one transition in time \( dt \) are infinitesimally of a higher order and neglected.
Thus the state transition equation can be developed using the above set of rules. The following diagram Fig.(2.1) helps in writing down the Markovian equation (Ryabinin [56], p.341).

![Markov graph for single component repairable system](image)

\[ \lambda_0 \, dt \] represents the probability of transition from state '0' to '1' and \( \mu \, dt \) represents the probability of transition from state '1' to '0' in time \( dt \). The system at any arbitrary moment of time 't' can be found to be either in state '0' with probability of transition equal to
Thus the estimate of probabilities $P_0(t + dt)$ and $P_1(t + dt)$ that the system is in state '0' and '1' at time $(t + dt)$ can be seen in the following ways respectively.

(i) For $P_0(t + dt)$, the system may be in state '0' at time 't', and sustains this state within the time interval $dt$ or

(ii) The system may be in state '1' at time 't' and is renewed (repaired within time interval $dt$) i.e., it passes into the state '0'. The two events represented above are disjoint and hence

$$\lambda_o \ dt = Q(dt) = 1 - \exp(-\lambda \ dt) \quad (2.2.3)$$

$$P_0(t + dt) = P_0(t) \left[ 1 - \lambda_o dt \right] + P_1(t) \mu dt$$

or $P'_0(t) = -\lambda_o P_0(t) + \mu P_1(t) \quad (2.2.4)$
The above set of equations can be solved using the initial conditions

\[ P_0(0) = 1, \quad P_1(0) = 0 \] (since it is assumed in reliability analysis that the system at time '0' is expected to be in good condition always) satisfying the requirement:

\[ P_0(t) + P_1(t) = 1, \quad \text{since system must be in one of the states say '0' or '1' at any time 't'.} \]

The system of equations can be solved using Laplace Transformation (LT) technique and the solution is

\[ P_0(t) = \left[ \mu (\lambda + \mu) \right] + \lambda \left[ \exp(-\lambda t) / (\lambda + \mu) \right] \]

(2.2.6)

\[ P_1(t) = \lambda \left[ 1 - \exp(-\lambda t) \right] / (\lambda + \mu) \]

(2.2.7)
and $P_0(t)$, $P_1(t)$ represent the probability that the system is operative and inoperative, at the instant of time 't' respectively. $P_0(t)$ probability that the system is working satisfactorily is termed as availability and denoted by $\pi(t)$ and is expressed (Ryabinin [56] p.342) as:

$$\pi(t) = P_0(t) = \left[1 + \eta \exp(-(\lambda + \mu)t)\right](1 + \eta)$$

(2.2.8)

where operating ratio

$$\eta = \lambda/\mu = \frac{T_r}{T_f}$$

$T_r$ = Expected time of repair

$T_f$ = Expected time of failures

Availability is thus a function of $\eta$ and $t$. Availability of the system is getting smaller for larger $\eta$ in Fig. (3.2) (as given by Ryabinin [56] p.343).
(i) In particular if $\eta = 0$, $\lambda = \text{constant}$

$\Rightarrow \mu = \infty$ or $T_r = 0$. This reveals a situation that the state of failure is instantaneously eliminated then $\pi(t) = 1$ which follows from (2.2.8).

![Diagram](image)

**FIG(2.2):** Influence of 'η' on the Availability Function $\pi(t)$.

(ii) If $\eta = \infty$, $\lambda = \text{constant}$

$\Rightarrow \mu = 0$ or $T_r = \infty$, this refers to the case of non-repairability, then

$$\pi(t) = R_s(t) = \exp (-\lambda t) \quad (2.2.9)$$
This represents the reliability of the non-repairable system (Probability of faultless operation).
This reveals that when service rate $\mu = 0$, availability $\pi(t)$ is obviously the reliability ($R_s(t)$ - fault-free operation).

(iii) If $\eta = \text{constant}$, $\lambda = \text{constant}$, then in this case there is a moment of time say '$t_s$' (moderately large) after which $\pi(t)$ is constant, which is known as steady-state availability. Therefore the steady state value of availability is

$$\pi(\infty) = \frac{1}{1 + \eta} \quad (2.2.10)$$

$\pi(\infty)$ is the measure of proportion of uptime of the system in the long run usage of the system. The probability that the system will be in good state at the moment '$t$' and further liable to operate faultlessly during the time '$t$' is expressed (Ryabinin [56] p.344) as:

$$R(t, \tau) = \pi(t) R(\tau)$$
if \( t \) is larger say \( t > t_s \) then the expression (2.2.11) becomes

\[
R(\infty, \tau) = \exp(-\lambda \tau)/(1 + \eta)
\]  
(2.2.12)

Thus it tells that renewal or repair in this problem improves the reliability of the system only in the sense that increasing the availability. In the absence of renewal (assuming \( \mu = 0 \)) (2.2.11) can be expressed as

\[
R(t, \tau) = \exp(-\lambda t) \exp(-\lambda \tau)
\]
\[
= \exp[-\lambda (t + \tau)]
\]  
(2.2.13)

\( K_e(t) \), the coefficient of efficiency of renewal is a quantitative measure for efficiency of renewal and is defined (Ryabinin [56] p.344) as the ratio of
the expressions (2.2.11) and (2.2.13). Thus

\[ K_e(t) = \frac{\exp(\lambda t) + \eta \exp(-\mu t)}{1 + \eta} \]  (2.2.14)

and for larger 't' i.e., \( t > t_s \) (2.2.14) can be written as

\[ K_e(\infty) = \frac{\exp(\lambda t)}{1 + \eta} \]

\[ = \mu \frac{\exp(\lambda t)}{(\lambda + \mu)} \]  (2.2.15)

The efficiency increases with the increase in intensity of renewal \( \mu \), the intensity of failure \( \lambda \) and time 't'.

2.3 RELIABILITY ANALYSIS FOR TWO-UNIT SYSTEM WITH REPAIR

The repair by consideration will improve the reliability of a parallel system. In a parallel system if one of the unit fails then the other parallel unit will continue to function and the one which failed will tried to be repaired. Parallel
system fails only when the second (parallel) unit fails before the one that is being restored to operation. The 2-component system can be represented by a three-state Markov model seen in the Fig.(2.3).

Fig.(2.3) Markov Graph for 2-unit system.

<table>
<thead>
<tr>
<th>State</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Both components are good.</td>
</tr>
<tr>
<td>1</td>
<td>One unit failed and the other working.</td>
</tr>
<tr>
<td>2</td>
<td>Both units failed.</td>
</tr>
</tbody>
</table>

For the reliability the instant of interest is the time of system reaching state 2 and what happens
after that is of no consequence. Therefore in the Markov graph there will not be a transition from state '2' to '1' and hence open (i.e., $\mu_2$ in the Fig.(2.3) is zero).

$\lambda_i, \mu_i$ are the rates of departure of the system from state $i$ to $i + 1$ and $i - 1$ respectively. As in the case of TWO unit system, $P_0(t + dt)$, $P_1(t + dt)$ and $P_2(t + dt)$ can be obtained and they are as follows (E. Balaguruswamy [12] p.123).

$$P_0(t + dt) = (1 - \lambda_0 dt) P_0(t) + P_1(t) \mu_1 dt \quad (2.3.1)$$

$$P_1(t + dt) = \lambda_0 dt P_0(t) + P_1(t) [1 - (\lambda_1 + \mu_1) dt] \quad (2.3.2)$$

$$P_2(t + dt) = \lambda_1 dt P_1(t) + P_2(t) \quad (2.3.3)$$

The following differential equations can be formulated.

$$P_0'(t) = -\lambda_0 P_0(t) + P_1(t) \mu_1 \quad (2.3.4)$$
\( P_1'(t) = -(\lambda_1 + \mu_1) P_1(t) + \lambda_0 P_0(t) \) \hspace{1cm} (2.3.5)

\( P_2'(t) = \lambda_1 P_1(t) \) \hspace{1cm} (2.3.6)

Using the Laplace transformation and using the initial conditions at \( t=0, P_0(0) = 1 \) and \( P_1(0) = P_2(0) = 0 \), the set of equations can be expressed

\[
\begin{bmatrix}
(s + \lambda_0) & -\mu_1 & 0 \\
-\lambda_0 & (s + \mu_1 + \lambda_1) & 0 \\
0 & -\lambda_1 & s
\end{bmatrix}
\begin{bmatrix}
P_0^*(s) \\
P_1^*(s) \\
P_2^*(s)
\end{bmatrix}
= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}
\]

(2.3.7)

and the solution vector of the above set of equations is
2.3.1 Reliability Function - Series Configuration

If the basic design of two-component system is series, the system is good when both the components are in good condition which is represented by the state '0'.

\[
P(s) = \begin{cases} 
\frac{[\lambda_1 + \mu_1 + s_1 \exp(s_1 t) - (\lambda_1 + \mu_1 + s_2) \exp(s_2 t)]}{(s_1 - s_2)} \\
\lambda_0 \frac{[\exp(s_1 t) - \exp(s_2 t)]}{(s_1 - s_2)} \\
1 + \frac{s_2 \exp(s_1 t) - s_1 \exp(s_2 t)}{(s_1 - s_2)} 
\end{cases}
\]

(2.3.8)

where

\[
s_1, s_2 = \frac{1}{2}\left[-(\lambda_0 + \lambda_1 + \mu_1) \pm \sqrt{(\lambda_0 + \lambda_1 + \mu_1)^2 - 4\lambda_0 \lambda_1^{1/2}}\right]
\]

(2.3.9)

Hence the probability that the system is in
state '0' represents the reliability of the system.

\[ R_s(t) = P_0(t) \]

where the parameters in \( P_0(t) \) are given by

\[ \lambda_0 = 2\lambda, \quad \lambda_1 = 0 \quad \text{and} \quad \mu_1 = 0 \]

This results in \( S_1 = 0, \quad S_2 = -2\lambda \), hence the reliability is expressed as

\[ R_s(t) = \exp(-2\lambda t) \quad \text{(2.3.10)} \]

### 2.3.2 Reliability Function of Parallel Configuration

For the parallel system to be good, the system must be either in state '0' or '1'. Thus the probability of the system successful operation is given by

\[ R_s(t) = P_0(t) + P_1(t) \]

\[ = 1 - P_2(t) \]
where parameters in $S_1$ and $S_2$ are defined in (2.3.9).

2.3.3 MTTF of System

For a system, mean time to system failure is an important characteristic which tells the average time the system is operative before it fails. This is obtained using the result in (1.3.3) and (2.3.11)

$$E_s(t) = \text{MTTF} = \int_0^\infty R_s(t)dt$$

where $R_s(t)$ is given in expression (2.3.11)

$$= -(S_1 + S_2) / S_1 S_2 \quad (2.3.12)$$

For a two-component system

$$S_1 + S_2 = -(\lambda_0 + \lambda_1 + \mu_1)$$

$$S_1 S_2 = \lambda_0 \lambda_1$$
and $\text{MTTF} = \frac{(\lambda_0 + \lambda_1 + \mu_1)}{\lambda_0 \lambda_1}$ \hspace{1cm} (2.3.13)

for the parallel system, it follows from (2.3.12) that

$$\text{MTTF} = \frac{3\lambda + \mu}{2\lambda^2}$$ \hspace{1cm} (2.3.14)

The expression of MTTF in the case of non-maintained system (i.e., $\mu = 0$) is seen from (2.3.13) as

$$\text{MTTF} = \frac{3}{2\lambda}$$ \hspace{1cm} (2.3.15)

Thus the percentage gain in MTTF of system due to maintenance is expressed (E. Balaguruswamy [12] p. 126) as

$$\text{GAIN} = \left[\frac{\text{MTTF} - \text{MTTF}|_{\mu=0}}{\text{MTTF}|_{\mu=0}}\right] \times 100$$

$$= 33.3(\mu/\lambda)$$ \hspace{1cm} (2.3.16)

in the case of parallel system. However in the case of series system no gain is realised.
2.4 SYSTEM AVAILABILITY ANALYSIS

The Markov approach for availability analysis is same as that of reliability analysis, but however, availability of the system is concerned with the status of the system at time 't' instead of probability of trouble-free operation of the system. That means the system would have been restored after it was found in down-state for sometime. In consequence, the repair at the state 2 of the Markov graph (see in Fig. (2.3)) is also considered in the case of availability. Thus the Markov graph for the availability analysis is as follows.

Fig. (2.4) Markov graph for availability of 2-component system.
Referring to the Markov graph, we could get the following set of equations which can be expressed in matrix notation using Laplace transformation and initial conditions i.e., \( P_0(0) = 1 \) and \( P_1(0) = P_2(0) = 0 \) (E. Balaguruswamy [12] p.126)

\[
\begin{pmatrix}
(s + \lambda_0) & -\mu_1 & 0 \\
-\lambda_0 & (s + \lambda_1 + \mu_1) & -\mu_2 \\
0 & -\lambda_1 & (s + \mu_2)
\end{pmatrix}
\begin{pmatrix}
P_0^*(s) \\
P_1^*(s) \\
P_2^*(s)
\end{pmatrix} =
\begin{pmatrix}
1 \\
0 \\
0
\end{pmatrix}
\]

(2.4.1)

We can solve the vector \( P^*(s) \) using the condition satisfying

\[
P_0^*(s) + P_1^*(s) + P_2^*(s) = 1/s
\]

or

\[
P_0(t) + P_1(t) + P_2(t) = 1
\]
and the solution is

\[ P^*(s) = \begin{bmatrix}
\frac{s^2 + s(\lambda_1 + \mu_1 + \mu_2) + \mu_1 \mu_2}{s(s-S_1)(s-S_2)} \\
\frac{s \lambda_0 + \lambda_0 \mu_2}{s(s-S_1)(s-S_2)} \\
\frac{1}{s} - (P^*_0(s) + P^*_1(s))
\end{bmatrix} \]

(2.4.2)

By taking inverse Laplace transformation of the terms given in the parentheses of (2.4.2) we get

\[ P_0(t) = \left[ \mu_1 \mu_2 / S_1 S_2 \right] + \left[ A_1 \exp(S_1 t) \right. \\
- A_2 \exp(S_2 t) \left. \right]/(S_1 - S_2) \]

(2.4.3)

where

\[ A_1 = \left( S_1^2 + S_1 (\mu_1 + \mu_2 + \lambda_1) + \mu_1 \mu_2 \right) / S_1 \]

\[ A_2 = \left( S_2^2 + S_2 (\mu_1 + \mu_2 + \lambda_1) + \mu_1 \mu_2 \right) / S_2 \]
and

\[ S_1, S_2 = (1/2)[-(\lambda_0 + \lambda_1 + \mu_1 + \mu_2) \pm \text{SQRT}((\lambda_0 + \lambda_1 + \mu_1 + \mu_2)^2 - 4(\lambda_0 \lambda_1 - \lambda_0 \mu_2 + \mu_1\mu_2)))] \]  

(2.4.4)

then

\[ P_1(t) = \left[ (\lambda_0/\mu_2)/S_1S_2 \right] + \left[ Z_1 \exp(S_1t) \right. \\
\left. - Z_2 \exp(S_2t) \right] / (S_1 - S_2) \]  

(2.4.5)

where

\[ Z_1 = S_1 \lambda_0 + \lambda_0 \mu_2 \]

\[ Z_2 = S_2 \lambda_0 + \lambda_0 \mu_2 \]

and \( S_1 \) and \( S_2 \) are given in (2.4.4).
2.4.1 Time Dependent Availability

(i) Series System

For the series system the successful operation of the system is represented only by the state '0' in the state space. Thus the probability of the successful operation of the system is nothing but \( P_0(t) \), thus

\[
\pi_s(t) = P_0(t) = \left[ \frac{\mu_1 \mu_2}{s_1 s_2} \right] + \left[ A_1 \exp(s_1 t) - A_2 \exp(s_2 t) \right] / (s_1 - s_2) \tag{2.4.6}
\]

and \( s_1 s_2 = \lambda_0 \lambda_1 + \lambda_0 \mu_2 + \mu_1 \mu_2 \)

\( A_1, A_2 \) are defined in (2.4.3) and depend on the basic model parameters \( \lambda_0, \lambda_1, \mu_1, \mu_2 \)

where

\[
\lambda_0 = 2\lambda, \quad \lambda_1 = \lambda, \quad \mu_1 = \mu_s, \quad \mu_2 = 2\mu_s \tag{\text{E. Balaguruswamy [12] p.127}} \tag{2.4.7}
\]
on the substitution of the above quantities in the expression (2.4.6).

Availability of the system as a function of \( t \) can become

\[
\pi_s(t) = \left[ \frac{2\mu_s^2}{(4\mu_s \lambda + 2\mu_s^2)} \right]_t [A_1 \exp(s_1 t) - A_2 \exp(s_2 t)] / (s_1 - s_2) \tag{2.4.8}
\]

(ii) Parallel System

In the case of parallel configuration, the system is operative if the system is represented by either in the state '0' or '1'. Thus the probability of successful operation of the system is indicated by

\[
\pi_p(t) = P_0(t) + P_1(t) \tag{2.4.9}
\]

where \( P_0(t) \) and \( P_1(t) \) are given by the expressions (2.4.3) and (2.4.5) and depend on basic parameter \( \lambda_0, \lambda_1, \mu_1 \) and \( \mu_2 \). By substituting them in the expressions respectively we get the expression of
availability in the case of parallel system. The values of the parameters in the case of parallel configuration are

\[ \lambda_0 = 2\lambda, \quad \lambda_1 = \lambda, \quad \mu_1 = \mu_s \quad \text{and} \quad \mu_2 = 2\mu_s \]

2.5 STEADY-STATE AVAILABILITY

(i) Series System

Steady-state availability of the system is nothing but the enquiry of the status of the system after long run usage of the system. In otherwords, it tells the performance of the system as \( t \to \infty \).

Thus the steady-state availability of the system is arrived at as a limiting case of the availability

\[
\pi(\infty) = \lim_{t \to \infty} \pi(t) = \lim_{t \to \infty} [P_0(t)]
\]

\[
= \lim_{s \to 0} [sp_0^*(s)],
\]
using final value theorem (Shooman [58] p.108)
where \( P_0^*(s) \) is the Laplace transformation of \( P_0(t) \).

Therefore

\[
\pi(\infty) = \frac{\mu_1 \mu_2}{S_1 S_2}
\]

\[
= \frac{\mu_1}{(\lambda_0 + \mu_1)}
\]

\[
= \frac{\mu}{(2\lambda + \mu)} \quad (2.5.1)
\]

Extending the result to \( n \)-component system the expression of steady-state availability for series system is

\[
\pi(\infty) = \frac{\mu}{(n \lambda + \mu)} \quad (2.5.2)
\]

(ii) Parallel System

Availability in the case of parallel system is

\[
\pi_p(t) = P_0(t) + P_1(t), \text{ following from (2.4.9)}.
\]
Therefore the steady-state availability of the parallel system is

\[ \pi_p(\infty) = \lim_{t \to \infty} \pi(t) \]

\[ = \lim_{t \to \infty} [p_0(t) + p_1(t)] \]

\[ = \lim_{s \to 0} [s p_0^*(s) + s p_1^*(s)] \]

using final value theorem

\[ \pi_p(\infty) = \frac{2\mu \lambda + \mu^2}{(\lambda^2 + 2\mu \lambda + \mu^2)} \quad (2.5.3) \]

Throughout, in this chapter, we briefly reviewed the Markov models and discussed the formulae regarding them, which can further be used in this thesis in latter chapters.