Review of Markov Models   Reference: Chapter 6 of Ebeling   Maghsoodloo

Definition. A collection of random variables, X_t, that are indexed by a parameter (most often time) are called a stochastic process. The word stochastic means pertaining to chance, or random. A stochastic process is characterized by its state space (or location set), R_x, and its index set that is denoted by T.

Example 10. (a) The growth of a population is more realistically modeled as a stochastic process, where X_t is the population size at time t. In this example, the state space, R_x, is discrete while the index set, time, is continuous. Other examples with discrete state space and continuous index set are the number of units in stock (or inventory level) at time t, and the number of passengers waiting in queue for a bus at any time of the day. (b) Suppose we have a product that is not denumerable (such as gasoline, cloth, hazardous waste, etc) but the inventory level is observed only at discrete epochs of time. Such a stochastic process has a continuous state space but a discrete index set T (or parameter space). (c) The content of a dam observed at any time t is an example of a stochastic process with both continuous state and parameter spaces. (d) Consider a component, such as a valve or a resistor, which is subject to failure but is inspected only one time per month (or perhaps per week, or even per day, but on an equal-interval basis). Upon inspection it is classified as satisfactory (0), unsatisfactory or derated or degraded (state 1) perhaps needing preventative maintenance, and failed (state 2). This represents an example of a stochastic process with discrete state (or location) space $R_x = \{0, 1, 2\}$ and discrete index set $T = \{0, 1, 2, 3, 4, \ldots\}$.

Definition. A stochastic process is said to be Markovian if the future behavior of the system, X_t, depends only on the present state, t_n, and not how the present state has been attained. That is, the pr of any future event, when the present state is known, is not at all altered by extra information about the past behavior of X_t. In order to translate this definition into a pr statement, we say that $X(t)$ is a Markov process iff

$$\Pr[X(t) \le x \mid X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots, X(t_0) = x_0] =$$

$$= \Pr[X(t) \le x \mid X(t_n) = x_n].$$

Whenever the parameter space, or the index set, T is discrete, we will call such a process a Markov chain, and as a result we have the classifications given in Table 3.

Table 3. The Four Possible Types of Markov Processes

State Space / Parameter Set	Discrete (Location)	Continuous (Location)
Discrete (Time)	**Markov Chain**	**Markov Chain**
Continuous (Time)	**Markov Process (With most applications to Reliability Engr)**	**Markov Process**

In the Example 10 part (d), suppose when the component is in state 0 (satisfactory), the pr that it will be still satisfactory at the next month's inspection is 0.65, i.e., $P_{00} = 0.65$. Further, the transition pr in one month from 0 to unsatisfactory state is 0.27, i.e., $P_{01} = 0.27$, and thus by necessity $P_{02} = 0.08$. Finally, the rest of the one- step transition probabilities are $P_{11} = 0.55$, $P_{12} = 0.45$, and $P_{22} = 1.0$. In the study of Markov chains, the one-step transition probabilities are summarized in the form of a one-step transition matrix P, which for this part (d) of Example10 is given below.

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0.65 & 0.27 & 0.08 \\ 0 & 0.55 & 0.45 \\ 0 & 0 & 1 \end{bmatrix} \end{array} \qquad (64)$$

The state 2 in the above one-step transition matrix is called absorbing, which implies when the component fails, the process is stopped for either repair (if it is repairable) or renewal with an item with identical characteristics, and then the same transition matrix will prevail the following ongoing process. The matrix P above is said to be stochastic because all P_{ij} values lie within the interval [0, 1], inclusive, and the sum of each row adds identically to 1. In addition, if the transition probabilities for one step stay (roughly) the same for all times, then the process is said to be stationary (or time-homogeneous). In the present example we are assuming that the same matrix P governs the transition prs in months 1, 2, 3, 4, 5, 6, …, (i.e., P is time-homogeneous or stationary because the transition pr from time 1 to time 2 is the same as from month 2 to 3, and so forth). A transition matrix P is said to be doubly-stochastic iff both its rows and columns add exactly to 1 (or 100%).

The notation $\mathbf{P}_{ij}^{(2)}$ represents the pr of transition from state i to j in two steps (or time units), and the notation $\mathbf{P}_{ij}^{(n)}$ represents the n-step transition pr from state i to state j, for n = 2, 3, 4, 5,…. For the above example, $\mathbf{P}_{00}^{(2)} = P_{00}\times P_{00} + P_{01}\times P_{10} + P_{02}\times P_{20} = 0.65\times 0.65 + 0.27\times 0 + 0.08\times 0 = 0.4225$; similarly, $\mathbf{P}_{01}^{(2)} = P_{00}\times P_{01} + P_{01}\times P_{11} + P_{02}\times P_{21} = 0.65\times 0.27 + 0.27\times 0.55 + 0.08\times 0 = 0.3240$; $\mathbf{P}_{02}^{(2)} = P_{00}\times P_{02} + P_{01}\times P_{12} + P_{02}\times P_{22} = 0.65\times 0.08 + 0.27\times 0.45 + 0.08\times 1 = 0.2535$. It can easily be verified that if we square the stochastic one-step transition matrix P, then we obtain the 2-step stochastic transition matrix, i.e.,

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{bmatrix} 0.65 & 0.27 & 0.08 \\ 0 & 0.55 & 0.45 \\ 0 & 0 & 1 \end{bmatrix}^2 = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0.4225 & 0.3240 & 0.2535 \\ 0 & 0.3025 & 0.6975 \\ 0 & 0 & 1 \end{bmatrix} \end{array}.$$

where the 1st row of the last matrix on the RHS is identical to the 2nd-step transition prs computed in the above paragraph. The above matrix shows that the pr of going from a satisfactory state 0 to the failed absorbing state 2 in two steps is $\mathbf{P}_{02}^{(2)} = 0.2535$. It can easily be verified, through induction, that the n-step transition pr for a time-homogeneous Markov chain is given by $P^{(n)} = \mathbf{P}^n$. For example, the 10-step transition matrix for the process (64) is given by

$$\mathbf{P}^{(10)} = \mathbf{P}^{10} = \begin{bmatrix} 0.65 & 0.27 & 0.08 \\ 0 & 0.55 & 0.45 \\ 0 & 0 & 1 \end{bmatrix}^{10} = \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 0.0135 & 0.0295 & 0.9570 \\ 0 & 0.0025 & 0.9975 \\ 0 & 0 & 1 \end{bmatrix} \end{array}.$$

In the study of Markov chains, we must 1st ascertain how many equivalence classes the chain has. In the matrix of equation (64), we say that state 2 is accessible from 0 and also from 1 but states 0 and 1 are not accessible from 2. Further, state 1 is accessible from 0, but state 0 is not accessible from

1 because $P_{10}^{(n)} = 0$ for all n. Hence, we say that states 0 and 1 do not communicate (because the transition $1 \rightarrow 0$ is impossible for any n) but only $0 \rightarrow 2$ and $1 \rightarrow 2$ simply because only $P_{i2}^{(n)} > 0$ for i = 0, 1, while $P_{2j} = 0$ for j = 0, 1, and n = 1, 2, 3, …. Thus, the chain in equation (64) has two equivalence classes: the location set {0, 1} is transient even if states 0 and 1 do not communicate, implying that the process eventually leaves these two locations with certainty, while state {2} is absorbing. In the Markov chain (64), if we assume that the failed item is either repairable in much less than one step (or replaceable with a new unit, or renewable with pr of 0.80), then our Markov matrix may take the following form.

$$\mathbf{P} = \begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{2} \end{array} \begin{array}{c} \begin{array}{ccc} \mathbf{0} & \mathbf{1} & \mathbf{2} \end{array} \\ \begin{bmatrix} \mathbf{0.65} & \mathbf{0.27} & \mathbf{0.08} \\ \mathbf{0} & \mathbf{0.55} & \mathbf{0.45} \\ \mathbf{0.80} & \mathbf{0.20} & \mathbf{0} \end{bmatrix} \end{array} \qquad \mathbf{(65)}$$

The chain in equation (65) now has only one equivalence class because every state communicates with all others (i.e., $i \leftrightarrow j$, for all i, j = 0, 1, 2). Note that although the 1-step transition pr from 1 to 0 is zero, the 2-step transfer probability $\mathbf{P_{10}^{(2)}} = 0.45 \times 0.80 = 0.36$, and the fact that $P_{01} = 0.27$ now shows that states 0 and 1 communicate (i.e., $0 \leftrightarrow 1$).

In the study of finite-state Markov chains, there are two possibilities: (1) exactly one equivalence class all of which communicate. Such a chain is said to be irreducible. Further, if such a chain is aperiodic, then it is called ergodic. (2) Markov chains with at least two equivalence classes. In case (2), the chain may contain transient, recurrent, and absorbing states. In reliability engineering the absorbing state is always the failed state of the system. In order to define these 3 types of equivalence classes, we need to define the recurrence time distribution. Definition. Let $\mathbf{f_{ii}^{(n)}} = P[X_n = i,\ X_r \neq i\ (r = 1, 2, \ldots, n-1) \mid X_0 = i]$; in other words, $\mathbf{f_{ii}^{(n)}}$ is the Pr that, starting initially in state i, the Markov chain returns to i for the very 1st time in n steps (i.e., it avoids state i until the nth step). Then $f_{ii}^{*} = \sum_{\mathbf{n=1}}^{\infty} \mathbf{f_{ii}^{(n)}}$ gives the pr that, starting initially in i, the process eventually returns to i. A state "i" is said to be recurrent iff, starting from i, the eventual return to i is certain, i.e., iff $f_{ii}^{*} = 1$; state i is transient iff $f_{ii}^{*} < 1$.

Now, consider a recurrent state i; then the mean recurrence time is given by $\mu_i = \sum_{n=1}^{\infty} n\, f_{ii}^{(n)}$, which is the average number of steps needed for the 1st return to i, having started in i. If μ_i is finite, then the location i is said to be positive recurrent; if $\mu_i = \infty$, then the state i is said to be null recurrent.

Thus the Markov matrix in (65) has one equivalence positive recurrent class. For example, $\mathbf{f_{11}^{(1)}} = 0.55$, $\mathbf{f_{11}^{(2)}} = 0.45\times0.20 = 0.09$, $\mathbf{f_{11}^{(3)}} = 0.45\times0.80\times0.27 = 0.0972$, $f_{11}^{(4)} = 0.45\times0.80\times0.65\times0.27 + 0.45\times0.80\times0.08\times0.20 = 0.06894$, etc. I am certain that if we compute more recurrence prs such as $f_{11}^{(5)}, f_{11}^{(6)}, f_{11}^{(7)}, \ldots$, then the sum $\sum_{n=1}^{\infty} \mathbf{f_{ii}^{(n)}} = 1$, showing that state 1 is recurrent. As a class property, since states 0 and 2 communicate with state 1, then states 0 and 2 are also recurrent.

__Definition.__ The period of a Markov chain is the number of eigenvalues of P with modulus (or length) equal to 1. For example, the period of the chain

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

is d = 3 because the three eigenvalues of P are 1, and $\dfrac{-1 \pm \sqrt{3}\, i}{2} = \dfrac{-1 \pm \sqrt{-3}\, i}{2}$, each of which is of unit modulus (or length).

In the study of ergodic Markov chains (ergodic means irreducible, positive recurrent, and aperiodic), the analyst is generally interested to determine what proportion of the times the system spends in state i, i.e., the limiting distribution of prs over all the m states of a Markov chain (regardless of the initial state). We define the limiting (or stationary, or long-term) distribution as

$$\Pi = \lim_{n \to \infty} \mathbf{P}^{(n)} = \lim_{n \to \infty} \mathbf{P}^{n} = \begin{bmatrix} \pi' \\ \pi' \\ \cdot \\ \cdot \\ \cdot \\ \pi' \end{bmatrix}$$

where $\boldsymbol{\pi}' = [\pi_0 \quad \pi_1 \quad \pi_2 \ldots \quad \pi_{m-2} \quad \pi_{m-1}]$ is a probability row vector such that $\sum_{i=0}^{m-1} \pi_i \equiv 1$. Because $\Pi = \lim_{n \to \infty} \mathbf{P}^{(n+1)} = \lim_{n \to \infty} \mathbf{P}^{n+1} = \left[\lim_{n \to \infty} \mathbf{P}^{(n)}\right]\mathbf{P} = \Pi\,\mathrm{P}$, then in order to solve for the steady-state transition prs, we must require that $\pi' = \pi'\mathbf{P}$, which implies that $\mathbf{P}'\boldsymbol{\pi} = \boldsymbol{\pi}$. Note that $\mathbf{P}'$ (the transpose of P) is not generally a stochastic matrix (unless P were doubly stochastic). We now illustrate the procedure by obtaining the steady-state prs for the Markov matrix of equation (65). We 1st write $\mathbf{P}'\boldsymbol{\pi} = \boldsymbol{\pi}$.

$$\begin{bmatrix} \mathbf{0.65} & \mathbf{0} & \mathbf{0.80} \\ \mathbf{0.27} & \mathbf{0.55} & \mathbf{0.20} \\ \mathbf{0.08} & \mathbf{0.45} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \end{bmatrix} = \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \end{bmatrix}$$

We now make the observation that $\mathbf{P}'\boldsymbol{\pi} = \boldsymbol{\pi}$ is the same system of equations as is $\mathbf{P}'\boldsymbol{\pi} - \boldsymbol{\pi} = (\mathbf{P}' - \mathbf{I}_3)\boldsymbol{\pi} =$ 0, where in this example I_3 is the identity matrix in D3. The homogeneous system $(\mathbf{P}' - \mathbf{I}_3)\boldsymbol{\pi} =$ 0 of 3 equations with 3 unknowns has no solution if the determinant of $(\mathbf{P}' - \mathbf{I}_3)$ is different from zero because the system will be inconsistent, and it has an infinite number of solutions iff det$((\mathbf{P}' - \mathbf{I}_3)) \equiv 0$. Since the det$((\mathbf{P}' - \mathbf{I}_3)) \equiv \begin{vmatrix} \mathbf{-.35} & \mathbf{0} & \mathbf{0.80} \\ \mathbf{0.27} & \mathbf{-.45} & \mathbf{0.20} \\ \mathbf{0.08} & \mathbf{0.45} & \mathbf{-1} \end{vmatrix} \equiv 0$, then the homogeneous system $(\mathbf{P}' - \mathbf{I}_3)\boldsymbol{\pi} =$ 0 has an infinite number of solutions. In order to obtain a unique solution we must impose the absolute required constraint that $\pi_0 + \pi_1 + \pi_2 \equiv 1$, and select any two of the 3 equations 0.65 π_0 + 0.80 π_2 = π_0, 0.27 π_0 + 0.55 π_1 + 0.20 π_2 = π_1, and 0.08 π_0 + 0.45 π_1 = π_2 plus the constraint $\pi_0 + \pi_1 + \pi_2 \equiv 1$ to solve simultaneously for the 3 unknowns. We arbitrarily select 0.65 π_0 + 0.80 π_2 = π_0, and 0.08 π_0 + 0.45 π_1 = π_2. Then our system of equations become

$$\begin{cases} -.35\pi_0 + \qquad + 0.80\pi_2 = 0 \\ 0.08\pi_0 + 0.45\pi_1 \quad - \pi_2 = 0 \\ \pi_0 + \quad \pi_1 \qquad + \pi_2 = 1 \end{cases} \rightarrow \quad A\,\pi = b, \qquad (66)$$

where the 3×3 matrix $A = \begin{bmatrix} -0.35 & 0 & 0.80 \\ 0.08 & 0.45 & -1 \\ 1 & 1 & 1 \end{bmatrix} = (\mathbf{P}' - \mathbf{I}_3)$, the 3×1 vector $\pi = \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \end{bmatrix}$, and the 3×1 column vector $b = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Equation (66) clearly shows that the unique solution $\pi = A^{-1}\,b =$

$$\begin{bmatrix} -1.8046 & -0.9956 & 0.4480 \\ 1.3441 & 1.4312 & 0.3559 \\ 0.4605 & -0.4356 & 0.1960 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.44804 \\ 0.35594 \\ 0.19602 \end{bmatrix}.$$

Therefore, the process on the average spends 44.8% of the times in satisfactory state, 35.6 % of times in degraded state, and 19.6% of times in the failed state. Note that if we raise the matrix P to the 20th power, we will practically obtain the 3×3 matrix Π , as shown below to 6 decimals (I used Matlab which provides 14-decimal accuracy).

$$P^{(20)} = P^{20} = \begin{bmatrix} 0.448040 & 0.355943 & 0.196017 \\ 0.448040 & 0.355943 & 0.196017 \\ 0.448040 & 0.355943 & 0.196017 \end{bmatrix} \cong \Pi$$

We now return to Markov chains which have at least two equivalence classes, such as the chain given in equation (64). We need this in reliability analysis because the failed state can be considered as absorbing. In such a chain the 1st step is to write the one-step transition matrix in canonical form, which lists absorbing states 1st, followed by recurrent classes, and last the transient classes. The canonical representation of process (64) is given by

$$\tilde{\mathbf{P}} = \begin{array}{c} \\ 2 \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} 2 & 0 & 1 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0.08 & 0.65 & 0.27 \\ 0.45 & 0 & 0.55 \end{bmatrix} \end{array} = \begin{bmatrix} \mathbf{P_1} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix} \quad \textbf{(67)}$$

where P_1 is in general an r×r submatrix with transition prs only among recurrent states (in this case r = 1), the (m – r)×(m – r) submatrix Q (the truncated matrix) provides the transition prs among transient states with at least one row whose sum must be less than 1, the (m – r)×r submatrix R gives the transition prs from transient to recurrent states, and 0 is an r×(m – r) submatrix all of whose elements are identically equal to zero. For our example, m = 3, Q is a 2×2 matrix of transitions among transient states, R is a 2×1 vector, and 0 is a 1×2 submatrix. Note that if a Markov chain starts in a recurrent class, then return to a transient class will be impossible so that we have to make the assumption that the process starts in a transient class. Then it follows that in many transitions we must have the expected number of visits from a transient state to transient states as given by the matrix

$$\mathbf{N} = 1\times\mathbf{I} + 1\times\mathbf{Q} + 1\times\mathbf{Q}^2 + 1\times\mathbf{Q}^3 + 1\times\mathbf{Q}^4 + \ldots = \sum_{n=0}^{\infty}\mathbf{Q}^n \quad \textbf{(68)}$$

Further after (n – 1) steps by simple multiplication we must have:

$$\mathbf{N}(\mathbf{I} - \mathbf{Q}) = (\mathbf{I} - \mathbf{Q})(\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \ldots + \mathbf{Q}^{n-1}) = \mathbf{I} - \mathbf{Q}^n \quad \textbf{(69)}$$

In equation (69) because the $\lim_{n \to \infty} \mathbf{Q}^n \equiv \mathbf{0}$, then it follows that in the long run (i.e., over many **time periods),** $\mathbf{N}(\mathbf{I} - \mathbf{Q}) = (\mathbf{I} - \mathbf{Q})(\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \ldots + \mathbf{Q}^{n-1}) \rightarrow \mathbf{I} \rightarrow \sum_{n=0}^{\infty}\mathbf{Q}^n = (\mathbf{I} - \mathbf{Q})^{-1}$. Thus, combining equations (68) and (69) we obtain the fundamental matrix

$$\mathbf{N} = \sum_{n=0}^{\infty}\mathbf{Q}^n = (\mathbf{I} - \mathbf{Q})^{-1} \quad \textbf{(70)}$$

The N_{ij} element of the matrix N gives the expected number of visits that (starting from a transient state i) the chain makes to the transient state j before finally leaving the transient states for a recurrent class. For the chain given in equation (67), the fundamental matrix is given by

$$\mathbf{N} = \sum_{n=0}^{\infty} \mathbf{Q}^n = (\mathbf{I}_2 - \mathbf{Q})^{-1} = \begin{bmatrix} 0.35 & -0.27 \\ 0 & 0.45 \end{bmatrix}^{-1} = \begin{matrix} & 0 & 1 \\ 0 & \left[2.8571 \right. & \left. 1.7143 \right] \\ 1 & \left[0 \right. & \left. 2.2222 \right] \end{matrix}.$$

The above fundamental matrix implies that if the component starts in the satisfactory state (0), then it will stay as satisfactory an average of 2.8571 months and it will stay 1.7143 months in the unsatisfactory (or derated) state before it eventually fails. However, if the process starts in the unsatisfactory (or degraded) state, then it will stay as unsatisfactory an average of 2.2222 months before it fails. The fundamental matrix, N, plays an essential role in RE analysis because it will be used to compute the $MTBF_{Sys}$ of a repairable system. It easily follows that the column vector (both i and j are transient states)

$$\mathbf{N_i} = \mathbf{N_C} = \begin{bmatrix} \sum_{j=1}^{m-r} N_{1j} \\ \sum_{j=1}^{m-r} N_{2j} \\ . \\ . \\ . \\ \sum_{j=1}^{m-r} N_{rj} \end{bmatrix} \quad \text{[The subscript C stands for summing over columns.]} \qquad (71)$$

gives the expected total number of visits the process makes to transient states before leaving for a recurrent state. Thus, if the component starts as satisfactory, on the average it takes 4.5714 months before it fails, while if it starts as unsatisfactory (or derated), then on the average it will take 2.2222 months before it fails, i.e., $N_C = \begin{bmatrix} 4.5714 \\ 2.2222 \end{bmatrix}$.

It can be proven (see U. Narayan Bhat (1984),“ *Elements of Applied Stochastic Processes*”, 2nd Edition, John wiley & sons Inc., pp. 71-82, ISBN:0-471-87826-X) that

$$V(N_{ij}) = N(2N_D - I) - N_2 \qquad (72)$$

where the matrix N_D is an (m – r)×(m – r) matrix with diagonal elements equal to those of N and zeros elsewhere, and N_2 is an (m – r)×(m – r) matrix whose elements are equal to the square of elements of N. Thus, for the example under consideration $N_D = \begin{bmatrix} \mathbf{2.8571} & \mathbf{0} \\ \mathbf{0} & \mathbf{2.2222} \end{bmatrix}$, and $N_2 =$

$$\begin{bmatrix} \mathbf{2.8571^2} & \mathbf{1.7143^2} \\ \mathbf{0^2} & \mathbf{2.2222^2} \end{bmatrix} = \begin{bmatrix} \mathbf{8.1633} & \mathbf{2.9388} \\ \mathbf{0^2} & \mathbf{4.9383} \end{bmatrix}$$

As a result, from equation (72) the variance matrix is given by

$$V(N_{ij}) = \begin{bmatrix} \mathbf{2.8571} & \mathbf{1.7143} \\ \mathbf{0} & \mathbf{2.2222} \end{bmatrix} \times \begin{bmatrix} \mathbf{4.7143} & \mathbf{0} \\ \mathbf{0} & \mathbf{3.4444} \end{bmatrix} - \begin{bmatrix} \mathbf{8.1633} & \mathbf{2.9388} \\ \mathbf{0^2} & \mathbf{4.9383} \end{bmatrix} =$$

$$= \begin{bmatrix} \mathbf{5.3061} & \mathbf{2.9660} \\ \mathbf{0} & \mathbf{2.7160} \end{bmatrix} \rightarrow V(N_i) = V(N_C) = \begin{bmatrix} \mathbf{8.2721} \\ \mathbf{2.7160} \end{bmatrix}.$$

The above developments imply that if the component starts in state 0, then the expected number of months that stays as satisfactory is 2.8571 with a standard deviation of $\sqrt{\mathbf{5.3061}}$ = 2.3035, and the expected number of months spent in the unsatisfactory state is 1.7143 with a standard deviation of $\sqrt{\mathbf{2.9660}}$ = 1.7222, etc. Further, having started as satisfactory, then on the average it will take 4.5714 months before it eventually fails with a variance $V(N_0)$ = 8.2721.

Next we are interested in the 1st passage probabilities from a transient state to a recurrent state. As an example, we would be interested in computing the pr that if the component starts in state 0 (or satisfactory) or state 1 (derated), what is the pr that it will fail (state 2) for the 1st time in 3 months. This Pr distribution for one step is given by the vector $R = \begin{bmatrix} \mathbf{0.08} \\ \mathbf{0.45} \end{bmatrix}$; then for two months the 1st passage pr vector is given by $F^{(2)} = Q \times R = \begin{bmatrix} \mathbf{0.65} & \mathbf{0.27} \\ \mathbf{0} & \mathbf{0.55} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.08} \\ \mathbf{0.45} \end{bmatrix} = \begin{bmatrix} \mathbf{0.1735} \\ \mathbf{0.2475} \end{bmatrix}$, and for 3-steps

$F^{(3)} = Q^2 \times R = Q(Q \times R) = Q \times F^{(2)} = \begin{bmatrix} \mathbf{0.65} & \mathbf{0.27} \\ \mathbf{0} & \mathbf{0.55} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.1735} \\ \mathbf{0.2475} \end{bmatrix} = \begin{bmatrix} \mathbf{0.179600} \\ \mathbf{0.136125} \end{bmatrix}$, and similarly

$F^{(4)} = Q^3 \times R = \begin{bmatrix} \mathbf{0.274625} & \mathbf{0.292275} \\ \mathbf{0} & \mathbf{0.166375} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.08} \\ \mathbf{0.45} \end{bmatrix} = \begin{bmatrix} \mathbf{0.15349375} \\ \mathbf{0.07486875} \end{bmatrix}$, etc. Note that the sum of

$F^{(n)}$ over all n in this example will have to add up to a pr vector whose elements must equal to 1 because the component will eventually have to fail and be absorbed in state 2. This is illustrated below.

$$F = \sum_{n=1}^{\infty} \mathbf{F}^{(n)} = F^{(1)} + \sum_{n=2}^{\infty} \mathbf{F}^{(n)} = R + \sum_{n=2}^{\infty} \mathbf{Q}^{(n-1)}\mathbf{R} = \sum_{n=1}^{\infty} \mathbf{Q}^{(n-1)}\mathbf{R} = [\sum_{n=1}^{\infty} \mathbf{Q}^{(n-1)}]\mathbf{R}$$

$= N \times R$; for our example, this becomes $F = \begin{bmatrix} \mathbf{2.8571} & \mathbf{1.7143} \\ \mathbf{0} & \mathbf{2.2222} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.08} \\ \mathbf{0.45} \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix}$.

Example 11. Consider a maintained system that contains three units in pure parallel redundancy (i.e., only one of the three energized units is needed for system success). The system is checked once every hour and if a failure is detected on any component, repair starts on it immediately. For the sake of simplicity, we assume that within one hour the pr that a unit fails, gets repaired, and then fails again within the same hour is practically zero. Thus, we have a Markov chain with one hour as one step and the number of working units as the states of the system, i.e., $R_x = \{0, 1, 2, 3\}$ and the index parameter is given by T = {0, 1, 2, 3, 4, ……}. By observing the system every hour for over a month, the following one-step transition matrix has been approximated.

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} \mathbf{0.15} & \mathbf{0.50} & \mathbf{0.25} & \mathbf{0.10} \\ \mathbf{0.02} & \mathbf{0.25} & \mathbf{0.55} & \mathbf{0.18} \\ \mathbf{0} & \mathbf{0.08} & \mathbf{0.70} & \mathbf{0.22} \\ \mathbf{0} & \mathbf{0.02} & \mathbf{0.l8} & \mathbf{0.80} \end{bmatrix} \end{array}$$

The above transition matrix tells us the pr that the system goes from 3 units working successfully to 2 operational units in one hour is $P_{32} = 0.18$, while $P_{20} = 0$, etc. Clearly all states communicate implying that the chain has only one equivalence class, and the fact that the 4 eigenvalues of P are equal to $\xi_1 = 1.00$, $\xi_2 = 0.5868646$, $\xi_3 = 0.2459071$, and $\xi_4 = 0.0672283$ shows that there is only one eigenvalue with modulus (or length) equal to 1 so that the period of the chain is equal to 1. Note that the period of an irreducible Markov chain is always equal to number of eigenvalues with length (or modulus) equal to 1. For example, the complex numbers $\frac{\sqrt{3}}{2} \pm \frac{i}{2}$ have length equal to 1. Further, since all 4 states are positive recurrent and the chain period is equal to 1 (i.e., aperiodic), then the chain is said to be ergodic. The fact that the smallest eigenvalue $\lambda_4 = 0.0672283$ is close to zero, then the chain approaches its steady-stare very rapidly. To obtain the limiting distribution, we have to solve the system of equations $\mathbf{P}'\pi = \pi$ for the pr vector π imposing the constraint $\pi_0 + \pi_1 + \pi_2 + \pi_3 \equiv 1$.

We first check on the determinant of $(\mathbf{P'} - \mathbf{I}_4)$. Matlab computations gives $|\mathbf{P'} - \mathbf{I}_4| \equiv 0$. We can now select any 3 of 4 equations from $\mathbf{P'}\boldsymbol{\pi} = \boldsymbol{\pi}$ with $\pi_0 + \pi_1 + \pi_2 + \pi_3 \equiv 1$ to obtain a unique simultaneous solution for π_0, π_1, π_2, and π_3. The resulting system of equations is

$$\begin{cases} \mathbf{0.15\pi_0 + 0.02\pi_1 = \pi_0} \\ \mathbf{0.50\pi_0 + 0.25\pi_1 + 0.08\pi_2 + 0.02\pi_3 = \pi_1} \\ \mathbf{0.25\pi_0 + 0.55\pi_1 + 0.70\pi_2 + 0.18\pi_3 = \pi_2} \\ \mathbf{\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1} \end{cases} \rightarrow \pi = \begin{bmatrix} \mathbf{0.001404} \\ \mathbf{0.059670} \\ \mathbf{0.421200} \\ \mathbf{0.517726} \end{bmatrix}$$

Thus, over one million hours, the system spends 1404 hours in state 0 (failed), 59670 hours with exactly one operational unit, 421200 hours with two operating units, and 517726 hours with all 3 units working successfully. Further, because the smallest eigenvalue, $\lambda_4 = 0.0672283$, is close to zero, in 8 transitional hours (from time 0) the system nearly reaches its steady-state as shown below.

$$P^{(8)} = P^8 = \begin{bmatrix} \mathbf{0.001495} & \mathbf{0.061561} & \mathbf{0.431560} & \mathbf{0.505384} \\ \mathbf{0.001465} & \mathbf{0.060977} & \mathbf{0.428451} & \mathbf{0.509108} \\ \mathbf{0.001452} & \mathbf{0.060731} & \mathbf{0.427132} & \mathbf{0.510685} \\ \mathbf{0.001357} & \mathbf{0.058651} & \mathbf{0.415511} & \mathbf{0.524480} \end{bmatrix}.$$

Suppose now that we wish to ascertain if we presently are, say in state 1, on the average how many hours it takes to end up in the 0 state, where all units are down and need repair (or replacement). To answer such a question, we first make the state zero as absorbing so that the chain will consist of 2 equivalence classes as shown below.

$$\tilde{\mathbf{P}} = \begin{matrix} & \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{0.02} & \mathbf{0.25} & \mathbf{0.55} & \mathbf{0.18} \\ \mathbf{2} & \mathbf{0} & \mathbf{0.08} & \mathbf{0.70} & \mathbf{0.22} \\ \mathbf{3} & \mathbf{0} & \mathbf{0.02} & \mathbf{0.l8} & \mathbf{0.80} \end{matrix} \qquad \mathbf{(73)}$$

The truncated matrix associated with the transition matrix $\tilde{\mathbf{P}}$ given in (73) is given as

$$\mathbf{Q} = \begin{bmatrix} 0.25 & 0.55 & 0.18 \\ 0.08 & 0.70 & 0.22 \\ 0.02 & 0.l8 & 0.80 \end{bmatrix}.$$

Note that Q must have at least one row whose Prs add to less than 1; otherwise something is wrong. As a result the fundamental matrix of the process is

$$\mathbf{N} = (\mathbf{I}_3 - \mathbf{Q})^{-1} = \begin{matrix} \\ 1 \\ 2 \\ 3 \end{matrix} \begin{matrix} \begin{matrix} 1 & \quad\quad 2 & \quad\quad 3 \end{matrix} \\ \begin{bmatrix} 50 & 349.0196 & 428.9216 \\ 50 & 358.8235 & 439.7059 \\ 50 & 357.8431 & 443.6275 \end{bmatrix} \end{matrix}$$

The above fundamental matrix shows that given an initial state 1, then on the average the system spends 50 hours in state 1, on the average it will spend 349.02 hours in state 2 (due to repair or renewal), and on the average it will spend 428.92 hours in state 3 before it finally fails. The vector $E(N_i) = N_C = \begin{bmatrix} 827.9412 \\ 848.5294 \\ 851.4706 \end{bmatrix}$ gives the expected (or mean) amount of total time spent by the system in states 1, 2, and 3, respectively, before eventually getting absorbed by state 0. We now use equation (72) to compute the variance of N_{ij} and then sum the elements in order to compute the variance of N_C.

$$\mathbf{V(N_{ij}) = N(2\,N_D - I_3) - N_2} = \begin{bmatrix} 2450 & 128310 & 196160 \\ 2450 & 128400 & 196350 \\ 2450 & 128400 & 196360 \end{bmatrix} \rightarrow \mathbf{V(N_C)} = \begin{bmatrix} 326920 \\ 327200 \\ 327210 \end{bmatrix}$$

Thus, given that the process starts in state 1, it will on the average stay in states 1, 2, or 3 a total of 827.9412 hours with a standard deviation of $\sqrt{\mathbf{326920}} =$ 571.7686 hours, and so forth.

The Mean Recurrence Time of a Markov Chain

Let $\pi = [\pi_0 \;\; \pi_1 \;\; \pi_2 \ldots \; \pi_{m-1}]' = [\pi_0 \;\; \pi_1 \;\; \pi_2 \ldots \; \pi_{m-1}]^T$ be the limiting distribution of an m-state ergodic Markov chain with the one-step transition pr matrix P. Let $\mathbf{T}_{ij}$ be the 1st passage time of the transition from state i to j, and let $\tau_{ij} = E(\mathbf{T}_{ij})$. Then it can be proven that the mean recurrence time

(MRT) is given by $\tau_{ii} = 1/\pi_i$. We illustrate this below (for more details see U. N. Bhat (1984), 2nd Ed., pp. 103-104). Suppose that the process starts in the state i at time 0. Then after one step, it will either locate in the state j, in which case the value of $\mathbf{T}_{ij}$ will equal to 1 with transition pr P_{ij}, or it will transfer to state $k \neq j$, in which case $\mathbf{T}_{ij} = 1 + \mathbf{T}_{kj}$ with transition pr P_{ik} ($k \neq j$). Hence the expected 1st passage time from i to j is given by

$$\tau_{ij} = E(\mathbf{T}_{ij}) = 1\times P_{ij} + \sum_{k\neq j} P_{ik} \times E(1 + \mathbf{T}_{kj}) = \sum_{k=0}^{m-1} P_{ik} + \sum_{k\neq j} P_{ik} \times E(\mathbf{T}_{kj})$$

$$= 1 + \sum_{k=0}^{m-1} P_{ik} \times E(\mathbf{T}_{kj}) - P_{ij}\, E(\mathbf{T}_{jj}) = 1 + \sum_{k=0}^{m-1} P_{ik} \times \tau_{kj} - P_{ij}\, \tau_{jj}. \qquad \textbf{(74)}$$

Let τ represent an m×m matrix whose elements give the expected 1st passage time from state i to state j, and whose diagonal elements τ_{jj} will represent the mean (or expected) recurrence time from j to the same state j. Since equation (74) gives the element of τ in the ith row and jth column, in matrix form we must have:

$$\tau = \mathbf{1} + P\times(\tau - \tau_D) \qquad \textbf{(75)}$$

where the matrix $\mathbf{1}$ is an m×m matrix all of whose elements are equal to 1, and τ_D is a diagonal matrix whose diagonal elements are equal to those of the m×m matrix τ. Premultipling equation (75) by the row vector $\pi' = \pi^T$ (the transpose of the column vector π) results in

$$\pi'\times\tau = \pi'\times\mathbf{1} + \pi'\times P\times(\tau - \tau_D) = \pi'\times\mathbf{1} + \pi'\times(\tau - \tau_D) \to 0 = \pi'\times\mathbf{1} + \pi'\times(-\tau_D) \to \pi'\times\tau_D =$$

$$\pi'\times\mathbf{1} \to \pi_i \times \tau_{ii} = 1 \to \tau_{ii} = 1/\pi_i \qquad \textbf{(76)}$$

Note that the same equation (76) would be arrived at if we 1st transpose equation (75), which gives $\tau' = \mathbf{1} + (\tau - \tau_D)'\times P'$, and 2nd post-multiply this resulting transposed matrix equation by the column vector π. The result will be $\tau_D'\times\pi = \mathbf{1}\times\pi$, which is not surprising because the transpose of $\pi'\times\tau_D = \pi'\times\mathbf{1}$ will yield the same as before.

Equation (76) shows that the MRT (mean recurrence time) for an ergodic Markov chain is simply given by $1/\pi_i$. For the example 11 above, we have $\tau_{00} = 1/0.001404 = 712.2507$ hours; this **implies as**

shown below. that given the system now is in state 0, then on the average it will take 712.2507 hours to return to state 0 for the very 1st time. Similarly, τ_{11} = 1/0.059670 = 16.7588 hours; τ_{22} = 1/0.421200 = 2.3742, and τ_{33} = 1/0.517726 = 1.9315 hours. Thus, if the system is in 3 operational state, then on the average it will take 1.9315 hours to return to state 3 for the very 1st time due to repair, etc. In order to compute the off-diagonal elements of the expected recurrence time matrix τ, the procedure is not as simple as obtaining the diagonal elements. If the ergodic chain has more than 5 states, hand calculations will become very tedious, and for m > 10, a set of codes has to be written in order to solve the system of equations. We 1st note that an inversion of τ is impossible.

$$\tau = \mathbf{1} + \mathbf{P}\times(\tau - \tau_D) \quad \rightarrow \quad (\mathbf{I_m} - \mathbf{P})\times\tau = \mathbf{1} + \mathbf{P}\times(-\tau_D) \quad \rightarrow$$

$\tau = (\mathbf{I_m} - \mathbf{P})^{-1}\times(\mathbf{1} - \mathbf{P}\times\tau_D)$ **;** However, since P is a stochastic matrix (i.e., the sum of each row is identically equal to 1), then the sum of each row of the matrix $(I_m - P)$ must identically equal to zero and hence the $\det(I_m - P) \equiv 0$, showing that $(I_m - P)^{-1}$ does not exist. Hence, we can solve the system of equations $\tau = \mathbf{1} + P\times(\tau - \tau_D)$ only through iterations. I will now illustrate the procedure for the Markov chain of Example 11. From equation (75), we have:

$$\tau = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 0.15 & 0.50 & 0.25 & 0.10 \\ 0.02 & 0.25 & 0.55 & 0.18 \\ 0 & 0.08 & 0.70 & 0.22 \\ 0 & 0.02 & 0.18 & 0.80 \end{bmatrix} \times \begin{bmatrix} 0 & \tau_{01} & \tau_{02} & \tau_{03} \\ \tau_{10} & 0 & \tau_{12} & \tau_{13} \\ \tau_{20} & \tau_{21} & 0 & \tau_{23} \\ \tau_{30} & \tau_{31} & \tau_{32} & 0 \end{bmatrix} \qquad \textbf{(77)}$$

The above system shows that τ_{00} = 1 + 0.50 τ_{10} + 0.25 τ_{20} + 0.10 τ_{30} = 712.2507; τ_{10} = 1 + 0.25 τ_{10} + 0.55 τ_{20} + 0.18 τ_{30}, and τ_{20} = 1 + 0.08 τ_{10} + 0.70 τ_{20} + 0.22 τ_{30}. Thus, the system of equations that yield the average recurrence time to state 0 is given by

$$\begin{cases} 0.50\,\tau_{10} + 0.25\tau_{20} + 0.10\,\tau_{30} = 711.2507 \\ -0.75\,\tau_{10} + 0.55\tau_{20} + 0.18\,\tau_{30} = -1 \\ 0.08\,\tau_{10} - 0.30\,\tau_{20} + 0.22\,\tau_{30} = -1 \end{cases}$$

→ τ_{10} = 827.9420 hours, τ_{20} = 848.5302, and τ_{30} = 851.4714 hours.

Thus, if the system starts in state 1, then on the average it will take 827.9420 hours to return to state 0 for the 1st time, etc. Similarly, from the matrix equation (77), we deduce that $\tau_{01} = 1 + 0.15\,\tau_{01} + 0.25\,\tau_{21} + 0.10\,\tau_{31}$, $\tau_{11} = 16.7588 = 1 + 0.02\,\tau_{01} + 0.55\,\tau_{21} + 0.18\,\tau_{31}$, and $\tau_{21} = 1 + 0.70\,\tau_{21} + 0.22\,\tau_{31}$. Solving this system of 3 equations with 3 unknowns yields $\tau_{01} = 10.00$ hours, $\tau_{21} = 20.5882$, and $\tau_{31} = 23.5294$ hours. The remaining mean recurrence times are $\tau_{02} = 3.3775$ hours, $\tau_{12} = 2.6879$, $\tau_{32} = 5.2689$, $\tau_{03} = 5.4071$ hours, $\tau_{13} = 4.8754$ and $\tau_{23} = 4.6334$ hours, which you may verify.

<u>Exercise 21.</u> Show that the Markov chain with the one-step transition Pr matrix

$$P = \begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} \end{array} \\ \begin{bmatrix} 0 & 0 & 0.60 & 0.40 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

has a period d = 3 and hence it is not ergodic. (b) A machine component is generally replaced as a rule every 5 months (as preventive maintenance). However, it has been found that this replacement policy is not adequate because the component sometimes wears out in less than 5 months. To ascertain whether to shorten the replacement cycle length, past history has shown that over many months, the lifetime distribution is as follows. Roughly 10% of components were replaced at the end of the 1st month; 15% of one-month old components were replaced at the end of the 2nd month; 35% of 2-month old components were replaced at the end of the 3rd month; and 40% of 3-month old components were replaced at the end of the 4th month. (i) Obtain the one-step transition pr matrix for the replacement policy. (ii) Obtain the age pr distribution for a component after the system has been in operation for a long time, where age = 0, 1, 2, 3, or 4 months. (iii) Obtain the expected replacement (or recurrence) time given that the system initially is in i = 0,1, 2, 3, 4 month old state. Answers: $\pi_0 = 0.288967$, $\pi_1 = 0.26007$, …, $\tau_{22} = 4.52366$ months, $\tau_{33} = 6.95948$, $\tau_{30} = 1.60$, $\tau_{10} = 2.734$, $\tau_{20} = 2.040$, $\tau_{23} = 2.87580$, $\tau_{13} = 4.24835$, and $\tau_{43} = 6.35943$ months.

Statistical Inference (SI) for Markov Chains

This is a difficult topic and thus I will restrict discussions only to point estimation for a 1st-order chain and test of hypothesis that a specific 1st-order transition matrix, P^0, may provide an adequate model for the observed (or realized) data. It will be best to explain SI through an example. Suppose that a system (such as a network) is observed once every day, and there are two possible states for the system. When the network is working then it is in state 0 and when it is down it is in state 1. Assume that the system is observed for a total of 50 days (this probably is not long enough time but will have to do in order to shorten the length of discussion). Initially we assume that the network is in state 0 (up). The 50-day observations have led to the following data : 0 0 0 0 1 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 0 0. Note that there were in fact 51 epochs of time because the system was observed initially to be in operating state, but there were 50 one-step transitions. The 50 transitions are summarized in Table 4. Thus the point ML estimates of P_{00} is given by $\hat{\mathbf{P}}_{00}$ = 32/39; $\hat{\mathbf{P}}_{01}$ = 7/39; $\hat{\mathbf{P}}_{10}$ = 7/11; $\hat{\mathbf{P}}_{11}$ = 4/11 so that the ML estimate of the transition matrix P is given by $\hat{\mathbf{P}}$ =

$$\begin{bmatrix} 32/39 & 7/39 \\ 7/11 & 4/11 \end{bmatrix}$$. The SI we would like to make is that "can we reasonably assume that the observed data is a realization of a Markov chain with the one-step transition matrix, say,

$P^0 \equiv \begin{bmatrix} 0.80 & 0.20 \\ 0.50 & 0.50 \end{bmatrix}$. Thus, we wish to test H_0 : $P = \begin{bmatrix} 0.80 & 0.20 \\ 0.50 & 0.50 \end{bmatrix}$ versus the

Table 4.

Initial State	**Final State**	**0**	**1**	$\mathbf{n_{i.}}$
0 (up)		**32 (31.2)**	**7 (7.8)**	**39**
1 (down)		**7 (5.5)**	**4 (5.5)**	**11**
$\mathbf{n_{.j}}$				**n = 50**

alternative H_1 : $P \neq P^0$ at the pre-assigned LOS $\alpha = 0.05$, i.e., we are willing to tolerate a 5% type I error. We use the χ^2 Goodness-of-Fit statistic to conduct this test, which is given by χ_0^2 =

$\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{(n_{ij}-E_{ij})^2}{E_{ij}}$, where n_{ij} is the observed frequency in the (ij) cell, and E_{ij} is the corresponding expected frequency computed under the null hypothesis. For our example, n_{11} = 32, n_{12} = 7, n_{21} = 7, and n_{22} = 4. In order to compute E_{11} we note that under H_0 the transition pr P_{00} is hypothesized to be 0.80, and since there are a total of $n_{1.}$= 39 transitions out of 0, then $E_{00} = 39\times P_{00} = 31.2$; however, we do not have another degree of freedom to compute E_{01} because $E_{00} + E_{01} = n_{1.} = 39$; this gives $E_{01} = 39 - 31.2 = 7.8$. Similarly, $E_{21} = 5.5$ and $E_{22} = 5.5$; these expected frequencies are provided in Table 6 in parentheses. Since our contingency Table 4 has 4 cells, then the net degrees of freedom (df) for our χ^2 Goodness-of-Fit statistic is $\nu = 4 -$ (the number of row constraints). Because the sum of each row prs in a Markov matrix must add to 1, then for a 2×2 Markov matrix we have 2 row constraints and hence our net df is equal to 4 – 2 = 2. The value of Chi-square statistic is $\boldsymbol{\chi_0^2} = \mathbf{\frac{(32-31.2)^2}{31.2}} + \mathbf{\frac{(7-7.8)^2}{7.8}} + \mathbf{\frac{(7-5.5)^2}{5.5}} + \mathbf{\frac{(4-5.5)^2}{5.5}} = 0.9207$. The critical value of $\boldsymbol{\chi_2^2}$ at the 5% LOS is given by the inverse function of $\boldsymbol{\chi_2^2}$ at a cumulative pr of 0.95, i.e., $\boldsymbol{\chi^2_{0.05;2}} = 5.9915$; thus, the data does not provide sufficient evidence to reject H_0 at the 5% level of significance. The pr level of the test (P-Value) is given by $\boldsymbol{\hat{\alpha} = P(\chi_2^2 \geq 0.9207)} = 0.6310 >> \alpha = 0.05$, as expected because the test statistic did not reject H_0.

It is worth noting that the χ^2 Goodness-of-Fit statistic is simply the weighted sum of squares of $\mathbf{\frac{(\hat{P}_{ij} - P_{ij}^0)}{\sqrt{P_{ij}^0}}}$ as shown below (this part is only fun reading that I conjured up!).

$$\boldsymbol{\chi_0^2} = \sum_{\mathbf{i=1}}^{\mathbf{m}}\sum_{\mathbf{j=1}}^{\mathbf{m}} \mathbf{\frac{(n_{ij}-E_{ij})^2}{E_{ij}}} = \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{(n_{ij}/n_{i.} - E_{ij}/n_{i.})^2}{E_{ij}/n_{i.}^2} = \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{n_{i.}(\hat{P}_{ij}-P_{ij}^0)^2}{E_{ij}/n_{i.}}$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{n_{i.}(\hat{P}_{ij}-P_{ij}^0)^2}{P_{ij}^0} = \sum_{i=1}^{m}\sum_{j=1}^{m} n_{i.}\left[\frac{\hat{P}_{ij}-P_{ij}^0}{\sqrt{P_{ij}^0}}\right]^2.$$

I hope that the reader will not be confused by the subscripts in this example because by cell (1, 1) we actually mean transition from state 0 to 0, and by cell (2, 1) we mean the transition from the system being down to being up in the next day. If we expand the expression for the $\chi_0^2 = \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{n_{i.}(\hat{P}_{ij}-P_{ij}^0)^2}{P_{ij}^0}$ we will have the following terms from rows 1 and 2:

$$\chi_0^2 = n_{1.}\left[\frac{(\hat{P}_{00}-P_{00}^0)^2}{P_{00}^0} + \frac{(\hat{P}_{01}-P_{01}^0)^2}{P_{01}^0}\right] + n_{2.}\left[\frac{(\hat{P}_{10}-P_{10}^0)^2}{P_{10}^0} + \frac{(\hat{P}_{11}-P_{11}^0)^2}{P_{11}^0}\right] \qquad \mathbf{(78)}$$

Substituting $\hat{P}_{01} = 1 - \hat{P}_{00}$ and $\hat{P}_{11} = 1 - \hat{P}_{10}$ into equation (78), we obtain:

$$\chi_0^2 = n_{1.}\left[\frac{(\hat{P}_{00} - P_{00}^0)^2}{P_{00}^0} + \frac{(\hat{P}_{00} - P_{00}^0)^2}{1 - P_{00}^0}\right] + n_{2.}\left[\frac{(\hat{P}_{10} - P_{10}^0)^2}{P_{10}^0} + \frac{(\hat{P}_{10} - P_{10}^0)^2}{1 - P_{10}^0}\right]$$

$$= n_{1.}(\hat{P}_{00} - P_{00}^0)^2\left[\frac{1}{P_{00}^0} + \frac{1}{1 - P_{00}^0}\right] + n_{2.}(\hat{P}_{10} - P_{10}^0)^2\left[\frac{1}{P_{10}^0} + \frac{1}{1 - P_{10}^0}\right]$$

$$= n_{1.}(\hat{P}_{00} - P_{00}^0)^2\left[\frac{1}{P_{00}^0(1 - P_{00}^0)}\right] + n_{2.}(\hat{P}_{10} - P_{10}^0)^2\left[\frac{1}{P_{10}^0(1 - P_{10}^0)}\right]$$

$$= \left[\frac{(\hat{P}_{00} - P_{00}^0)^2}{P_{00}^0(1 - P_{00}^0)/n_{1.}}\right] + \left[\frac{(\hat{P}_{10} - P_{10}^0)^2}{P_{10}^0(1 - P_{10}^0)/n_{2.}}\right] =$$

$$= \left[\frac{(\hat{P}_{00} - P_{00}^0)}{\sqrt{P_{00}^0(1 - P_{00}^0)/n_{1.}}}\right]^2 + \left[\frac{(\hat{P}_{10} - P_{10}^0)}{\sqrt{P_{10}^0(1 - P_{10}^0)/n_{2.}}}\right]^2 = \mathbf{Z_1^2} + \mathbf{Z_2^2} = \chi_1^2 + \chi_1^2 = \chi_2^2$$

The above developments, again, clearly show that the SMD of the statistic $\sum_{\mathbf{i=1}}^{\mathbf{m=2}}\sum_{\mathbf{j=1}}^{\mathbf{2}}\frac{\mathbf{n_{i.}(\hat{P}_{ij} - P_{ij}^0)^2}}{\mathbf{P_{ij}^0}}$ follows a Chi-square with 2 degrees of freedom. For a 3×3 Markov chain the df is equal to m(m – 1) = 6.

Exercise 22. A reliability system contains 2 components in pure parallel redundancy. Let X_n = The number of working components at time n, n = 0, 1, 2, 3 hours, …. and R_x ={0, 1, 2}. The system has been observed every hour, a total of 81 times with the following results. Initially the system was in state 2 (i.e., both units operating reliably) as shown by the data. 2212200122; 2211121222; 2110001211; 1222212210; 2002121122; 2212112222; 0221112221; 1101010122; 1
(a) Obtain the MLE (maximum likelihood estimates) of P_{ij} (i, j = 0, 1, 2). (b) Test the null hypothesis that $P = \begin{bmatrix} \mathbf{0.2} & \mathbf{0.30} & \mathbf{0.50} \\ \mathbf{0.40} & \mathbf{0.40} & \mathbf{0.20} \\ \mathbf{0.25} & \mathbf{0.40} & \mathbf{0.35} \end{bmatrix}$. Note that X_n could have just as well been defined as the number of failed units.

Markov Chains With Countably Infinite States

There are many stochastic processes where the states of the system can grow indefinitely such that the state space of the process is given by R_x = {0, 1, 2, 3, 4, ….}. One example, is a simple queueing system where the arrival rate (or birth rate), λ, is larger than departure rate (or death rate, or service rate) and the waiting room is unlimited. Under such a condition external factors have to be brought in to control the queue length (such as turning some customers away, or speeding up service so that λ_r = service rate $> \lambda$ = arrival rate, or add another server). In general, for a stochastic process if the traffic intensity (or utilization factor) $\rho = \lambda/\mu = \dfrac{\textbf{arrival rate}}{\textbf{service rate}} > 1$, then the birth-and-death process will grow unbounded. Thus, for a system to be stable, we must require that $0 < \rho = \lambda/\mu < 1$; If we allow the case $\rho = \lambda/\lambda_r = 1$, then the average recurrence time to state zero (i.e., empty system) may become infinite. To illustrate the concepts for a denumerable state simple birth-death (read as birth & death) process, consider a small shop with only one mechanic who adopts the policy of accepting new repair jobs on at most 2 cars (in queue) while he is working on one. Let X_n represent the number of jobs actually waiting in queue excluding the one under repair, i.e., state 0 represent the case where either one car is under repair and none is waiting , or also the facility is completely empty. For the sake of simplicity, we assume that the arrival distribution during one service period is given by

Number of arrivals	**0**	**1**	**2 or more**
Probability	**d = 0.15**	**0.75**	**b = 0.10**

The transition Pr matrix can be obtained from the arguments: P_{00} = Pr(0 or exactly one new job arrival) = 0.15 + 0.75 = 0.90; P_{01} = Pr(2 or more job arrivals) = 0.10. This leads to

$$
\mathbf{P} = \begin{array}{c} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \\ \\ \\ \\ \end{array}
\begin{array}{c}
\begin{array}{cccccccc} \mathbf{0} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & & \end{array} \\
\left[\begin{array}{cccccccc}
\mathbf{0.9} & \mathbf{0.1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0 \ldots} & & \\
\mathbf{0.15} & \mathbf{0.75} & \mathbf{0.1} & \mathbf{0} & \mathbf{0} & \mathbf{0 \ldots} & & \\
\mathbf{0} & \mathbf{0.15} & \mathbf{0.75} & \mathbf{0.1} & \mathbf{0} & \mathbf{0 \ldots} & & \\
\mathbf{0} & \mathbf{0} & \mathbf{0.15} & \mathbf{0.75} & \mathbf{0.1} & \mathbf{0} & \mathbf{0 \ldots} & \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0.15} & \mathbf{0.75} & \mathbf{0.1} & \mathbf{0} & \mathbf{0 \ldots} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & &
\end{array}\right]
\end{array}.
$$

Our primary objective is to determine what proportion of the times there are no jobs waiting for repair, i.e., we wish to compute π_0, which is also the proportion of the times the queue length is zero. To this end, we define a pr generating function (pgf) as $G(z) = \sum_{n=0}^{\infty} \pi_n z^n$, $|z| < 1$. The constraint $|z| < 1$ dictates that all the roots in the denominator of $G(z)$ must lie outside of a unit circle, or else $G(z)$ will not qualify to be a pgf (you will observe this later). Note that $G(z)$ is a pgf because $G(z = 0) = \pi_0$; $\left.\frac{dG(z)}{dz}\right|_{z=0} = \left.\sum_{n=1}^{\infty} n\pi_n z^{n-1}\right|_{z=0} = \pi_1$; $\left.\frac{d^2G(z)}{dz^2}\right|_{z=0} = \left.\sum_{n=2}^{\infty} n(n-1)\pi_n z^{n-2}\right|_{z=0} = 2!\ \pi_2$; similarly, $\left.\frac{d^3G(z)}{dz^3}\right|_{z=0} = 3!\ \pi_3$, etc. In other words, $G(z)$ generates probabilities through different powers of z, and it is constrained to the fact that $G(\text{at } z = 1) \equiv \sum_{n=0}^{\infty} \pi_n (1)^n = 1$. In order to solve for $G(z)$ we use $\pi' P = \pi'$, where $\pi' = [\pi_0 \quad \pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4 \ \ldots..]$ is a $1\times\infty$ row vector. The 1st four of these infinite equations with infinite unknowns are given below:

$$0.9\,\pi_0 + 0.15\,\pi_1 = \pi_0$$
$$0.1\,\pi_0 + 0.75\,\pi_1 + 0.15\,\pi_2 = \pi_1$$
$$0.1\,\pi_1 + 0.75\,\pi_2 + 0.15\,\pi_3 = \pi_2$$
$$0.1\,\pi_2 + 0.75\,\pi_3 + 0.15\,\pi_4 = \pi_3, \text{ etc.}$$

The above system can be solved recursively, but we choose a more elegant procedure that will be more general and will work when recursion is impossible. We multiply the above 4 equations by the proper powers of z as dictated by the RHS.

$$(0.9\,\pi_0 + 0.15\,\pi_1 = \pi_0)\, z^0$$
$$(0.1\,\pi_0 + 0.75\,\pi_1 + 0.15\,\pi_2 = \pi_1)\, z^1$$
$$(0.1\,\pi_1 + 0.75\,\pi_2 + 0.15\,\pi_3 = \pi_2)\, z^2$$
$$(0.1\,\pi_2 + 0.75\,\pi_3 + 0.15\,\pi_4 = \pi_3)\, z^3,$$

etc., which upon summing both sides reduces into the following equation:

$0.9\,\pi_0 + 0.15\,(\pi_1 + \pi_2 z + \pi_3 z^2 + \pi_4 z^3 + \ldots) + 0.1\,(\pi_0 z + \pi_1 z^2 + \pi_2 z^3 + \pi_3 z^4 + \ldots)$

$+ 0.75\,(\pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \pi_4 z^4 + \ldots) = G(z) \rightarrow$

$$0.9\,\pi_0 + \frac{0.15}{z}(\pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \pi_4 z^4 + \ldots) + 0.1\,z\,(\pi_0 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \ldots)$$

$+ 0.75\,[G(z) - \pi_0] = G(z) \rightarrow$

$$0.9\,\pi_0 + \frac{0.15}{z}[G(z) - \pi_0] + 0.1\,z\,G(z) + 0.75\,[G(z) - \pi_0] = G(z)$$

To simplify this last equation we multiply thru by z and collect terms of the same powers of z:

$$0.9\,\pi_0 z + 0.15\,[G(z) - \pi_0] + 0.1\,z^2 G(z) + 0.75z\,[G(z) - \pi_0] = z\,G(z) \rightarrow$$

$0.10\,z^2 G(z) - 0.25\,z\,G(z) + 0.15\,G(z) = 0.15\,\pi_0 - 0.15\,\pi_0 z \rightarrow$ Thus,

$$G(z) = \frac{0.15\pi_0(1-z)}{0.10z^2 - 0.25z + 0.15} = \frac{15\pi_0(1-z)}{10z^2 - 25z + 15} = \frac{3\pi_0(1-z)}{2z^2 - 5z + 3} =$$

$$= \frac{3\pi_0(1-z)}{(1-z)(3-2z)} = \frac{3\pi_0}{3-2z} = \frac{\pi_0}{1-2z/3}$$; Since $G(z = 1) = 1$, then it follows that $1 =$

$$\frac{\pi_0}{1-2/3} \rightarrow \pi_0 = 1/3 \rightarrow G(z) = \frac{1/3}{1-2z/3} = \frac{1}{3}\left[1 + \frac{2}{3}z + \left(\frac{2}{3}\right)^2 z^2 + \ldots\right]$$

$$\rightarrow \pi_1 = \frac{1}{3}\left(\frac{2}{3}\right),\ \pi_2 = \frac{1}{3}\left(\frac{2}{3}\right)^2,\ \pi_3 = \frac{1}{3}\left(\frac{2}{3}\right)^3 \ldots \rightarrow \pi_n = \frac{1}{3}\left(\frac{2}{3}\right)^n,\ n = 0, 1, 2, 3, 4, \ldots$$

Therefore, in the long run the proportion of times that the number of jobs waiting is zero (i.e., the queue length is zero) is 1/3, and the proportion of the times that the queue length is 1 is equal to 2/9; the proportion of the times that the queue length is 2 is equal to 4/27, etc. Note that in the above example, if the transition pr P_{01} (= pr 2 or more arrivals = 0.10 = b) were larger than P_{10}(= 0.15 = d), then the above system would become unstable. This is due to the fact that for a time-stationary simple birth-death process with a birth pr in a single period equaling b and a death pr equaling d, it can be shown that in general

$$\pi_n = \pi_0 \left(\frac{b}{d}\right)^n,\ n = 0, 1, 2, 3, 4, \ldots. \qquad (79)$$

In equation (79) when $b \geq d$, then the queue length (or the population size) will grow unboundedly (i.e., Lim π_n as $n \rightarrow \infty$ will tend to infinity). In fact the reader may verify thru the same procedure as I illustrated above that if we let b = 0.30 and d = 0.10, then the pgf will become $G(z) = \pi_0 / (1 - 3z)$

and $\pi_n = \frac{1}{3}(3)^n$. Note immediately that $G(z) = \frac{\pi_0}{1-3z}$ does not even qualify as a pgf because

$$G(1) = \sum_{n=0}^{\infty} \pi_n (1)^n = \sum_{n=0}^{\infty} \pi_n = -\pi_0/2 < 0,$$ which is impossible because the $\sum_{n=0}^{\infty} \pi_n \equiv 1$.

Furthermore, the reader should observe that the simple (or 1st-order) pole of $G(z) = \frac{\pi_0}{1-3z}$ is given by z = 1/3 which is inside a circle of radius 1 and not permissible. Finally, $G(z) = \pi_0 [1 + 3z + 9z^2 + (3z)^3 + 81z^4 + \ldots]$ shows that the queue length grows unlimitedly.

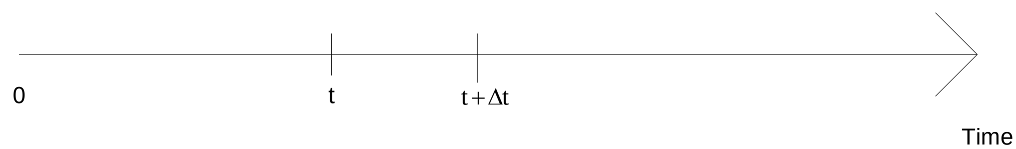
Markov Processes with Discrete State Space (or Location Set) R_x and Continuous Index Set T (Applications to RE Engineering)

From a reliability engineering standpoint a continuous index parameter Markov process amounts to practically observing the system on a continual basis (i.e., not at discrete epochs of time). We lay down the foundation by first discussing the general birth-death model on a continuous time scale, where X(t) represents the number of units failing by time t measured from zero (i.e., measured from the instant that the system has been put in service). In general, X(t) may represent the number of events (such as failures, or arrivals) occurring in the interval [0, t]. Please note that some authors refer to such a stochastic process as a chain because the state space is discrete, but we will use the terminology Markov process when the index parameter T is continuous. For starters, we must recognize that no longer we will be dealing with one-step, 2-step (etc) transition matrix because time is continuous and our one-step from a chain now changes to an infinitesimal length of time Δt for a continuous-time Markov process. Furthermore, since we will eventually take the limit as $\Delta t \to 0$, then we will assume that Δt is sufficiently small such that it will be impossible for two events to occur within the infinitesimal time interval $[t, t + \Delta t]$. Furthermore, $(\Delta t)^2 << \Delta t$ so that all terms containing $(\Delta t)^2$ will be ignored. We will now state the assumptions of a B-D (birth & death, or failure & repair, arrival & departure) process as follows:

(i) Birth: When the population size at time t is n, then the birth rate is given by λ_n (= instantaneous occurrence pr per time unit, i.e., $\lambda_n = \underset{\Delta t \to 0}{\text{Limit}} \frac{P_{n,n+1}(\Delta t)}{\Delta t}$). This implies that the pr of a single birth or failure during [t, t + Δt] is given by $\lambda_n \Delta t$, and because we are not allowing two or more failures during any interval of length Δt, then the Pr of no birth during [t, t + Δt] is given by $1 - \lambda_n \Delta t$. That is, $P[X(t+\Delta t) = n + 1 \mid X(t) = n] = \lambda_n \Delta t$, and $P[X(t+\Delta t) = n \mid X(t) = n] = 1 - \lambda_n \Delta t$. Further, births (or failures) occurring during [t, t + Δt] are independent of when the last birth occurred (the memoryless or Markovian property).

(ii) Death: Given that the population size at time t is n (> 0), then the infinitesimal Pr of death (or repair, or departure) is given by $r_n\Delta t$, and the Pr that a repair is not completed in the interval [t, t + Δt] is given by $1 - r_n\Delta t$. The value of Δt is sufficiently small such that the occurrence Pr of two or more repairs (or departures) during any interval of length Δt is zero. The death rate, r_n, when the population size is n (> 0), is the infinitesimal death-rate per unit of time. Further, repairs occurring during [t, t + Δt] are independent of when the last death occurred. The Pr of a death when X(t) = 0 is zero. Finally, birth and death of components (or arrivals/departures) in a system occur independently of each other.

We are now in a position to write the Chapman-Kolmogorov equations for the time intervals depicted below.



$$\mathbf{P_n(t + \Delta t) = P_n(t)\times[1 - \lambda_n\Delta t]\times[1 - r_n \Delta t] + P_{n-1}(t)[\lambda_{n-1} \Delta t]\times[1 - r_{n-1}\Delta t] +}$$
$$\mathbf{+ P_{n+1}(t)\,[r_{n+1} \Delta t] \times [1 - \lambda_{n+1} \Delta t]} \qquad \mathbf{(80a)}$$

In other words, the above equation (80a) tells us that the Pr of finding n individuals (or n failed units) in the system at time t + Δt (the LHS) is equal to the Pr of finding the system in state n and no birth or death during [t, t + Δt] , plus the Pr of system being in state n – 1 (failed units) followed by exactly one birth during [t, t + Δt] and no death during [t, t + Δt] , plus the mutually-exclusive Pr of process being in the state n+1 (failed units) followed by exactly one death (or one repair) and no birth during the interval [t, t + Δt]. Since products such as $\lambda_n\Delta t\times r_n\Delta t = \lambda_n r_n\times(\Delta t)^2$ are assumed to be practically zero relative to $\lambda_n \Delta t$ (or $r_n \Delta t$), equation (80a) reduces to

$$\mathbf{P_n (t + \Delta t) = P_n(t)\times[\, 1 - \lambda_n \,\Delta t - r_n \,\Delta t] + P_{n-1}(t)(\lambda_{n-1}\,\Delta t) + P_{n+1}(t)\times(r_{n+1}\,\Delta t)}$$

Transposing $P_n(t)$ to the LHS, dividing by Δt, and then taking the limit as $\Delta t \to 0$, this last equation becomes

$$\mathbf{\lim_{\Delta t \to 0} \frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = P_n(t)\times[\, -\lambda_n \; - r_n] + \lambda_{n-1}\, P_{n-1}(t) \; + r_{n+1}\, P_{n+1}(t) \;\rightarrow}$$

$$\mathbf{\frac{dP_n(t)}{dt} = P'_n(t) = -\,(\lambda_n + r_n\,)P_n(t) + \lambda_{n-1}P_{n-1}(t) \; + r_{n+1}P_{n+1}(t)} \qquad \textbf{(80b)}$$

$$\mathbf{P'_0(t) = -\,\lambda_0\, P_0(t) \; + r_1\, P_1(t)} \qquad \textbf{(80c)}$$

The relations in (80a & b) are called the difference differential (Chapman-Kolmogorov) equations (ddes) with the initial conditions: $P_n(0) = \begin{cases} \mathbf{1,\ \ if\ n = n_0} \\ \mathbf{0,\ \ if\ n \neq n_0} \end{cases}$, i.e., we assume that the systems starts initially with $n_0 \geq 0$ units (or generators) at time zero, and the Pr of finding the system in any other state but n_0 at time 0 is zero. Solving the above system of equations for general n > 2 is a formidable task and is well way beyond the objectives of this course. The most general solution will have, I am sure, a very terribly ugly expression, which I have not been able to find in the literature when $\lambda_n \neq \lambda_m$ and $r_n \neq r_m$ for all $n \neq m$. Note that this most general case pertains to the situation that the transition rates λ_n and r_n out of state n are actually state dependent and hence very difficult to solve. I do not know where this most general solution is provided in the literature (I am fairly sure the solution is probably given somewhere). The special case when λ_n and r_n are not state dependent (i.e., the stationary case) and are constant such that $\lambda_n = \lambda$ (for all n = 0, 1, 2, 3, 4, ….) and $r_n = \lambda_r$ (for all n = 1, 2, 3, 4, …) and (80b) reduces to

$$\mathbf{\frac{dP_n(t)}{dt} = P'_n(t) = -\,(\lambda + \;\lambda_r)\, P_n(t) + \lambda\, P_{n-1}(t) \; + \lambda_r\, P_{n+1}(t)} \qquad \textbf{(80d)}$$

$$\mathbf{P'_0(t) = -\,\lambda\, P_0(t) \; + \lambda_r\, P_1(t)} \qquad \textbf{(80e)}$$

The system of ddes in (80 d &e) assumes a single-channel Poisson input (i.e., number of failures occurring is Poisson distributed) and service time (or restoration time, or death-rate) is exponentially distributed. In Queueing theory this is referred to as the Markovian queue M/M/1, i.e., (M for Markov), where the interarrival times and services times are exponentially distributed with a single server. Again the time-dependent (or transient solutions) for the system of dde in (80 d &e) are very difficult to obtain and will involve the sum of modified Bessel functions. [The curious soul is referred to the "*Elements of*

Queueing Theory with Applications, by T. L. Saaty, Office of Naval Research, McGraw-Hill Book Company, ISBN: 54370, pp. 88-96"; or, "*The Theory of Stochastic Processes,* by Cox and Miller, Wiley Publications, 1968, pp. 192-196; or " *Queueing Systems, Volume 1*: *Theory,* by L. Kleinrock, John Wiley & sons, Inc., pp. 53-78; the solutions provided by Saaty in his equation (4-24) atop page 93 and by Kleinrock at the bottom of his page 77 are easier to swallow than any other solution I have seen in the literature, where their i = n_0 = the number of generators at initial time zero, but these solutions seem to contain an infinite sum of Bessel functions instead of a finite sum as reported by others in the literature; I do not understand the discrepancies. The modified 1st-kind Bessel function of order k is given by $I_k(x) = \sum_{j=0}^{\infty} \frac{(x/2)^{2j+k}}{j!\,(j+k)!}$; I hope this definition will help in understanding the general solution provided by the authors Saaty and Kleinrock.]. Just to satisfy the reader's curiosity, I will provide the general transient solution to the system of ddes (80 d &e) given by Leonard Kleinrock at the bottom of his page 77 (with his equation number 2.163).

$$\mathbf{P_n(t) = e^{-(\lambda+\mu)t}} \Big[\rho^{(n-n_0)/2}\, I_{n-n_0}(a\,t) \;+\; \rho^{(n-n_0-1)/2}\, I_{n+n_0+1}(a\,t) \;+$$

$$+\,\mathbf{(1-\rho)\rho^n} \sum_{j=n+n_0+2}^{\infty} [\rho^{-j/2} I_j(a\,t)] \Big] \qquad \mathbf{(81)}$$

where the traffic intensity $\rho = \lambda/\mu$, $a = 2\,\mu\,\rho^{1/2}$, and $I_k(x)$ is the modified 1st-kind Bessel function of order k defined above. I hope that I have satisfied your curiosity about the complexity of obtaining general solutions to any transient (i.e., time- dependent) stochastic process. Eq. (81) of Kleinrock provides the time-dependent solution to the most elementary queueing system, namely M/M/1. Note that when (81) represents the solution to a queueing system, then state n represents the total number of customers in the system including the one who is being served.

The reader should now be cognizant of the fact that we will not try to solve the difference differential equations (ddes) in (80) for the most general case, but I will try to solve them for the more special cases that arise in real-life reliability engineering. The good news in regard to applications to Reliability engineering is that in almost all cases the state space (or location set) in reliability models is finite (not infinite). Thus I will take up the cases one by one, but before we continue to take up the special cases, we would like to make a remarkable observation about (80 b&c) so that in future you could almost write down these ddes in matter of 2 minutes for a specific system. To do so, we concentrate only on nodes n (i.e., the system is either in state n items under repair), n – 1, or n+1 as depicted below.

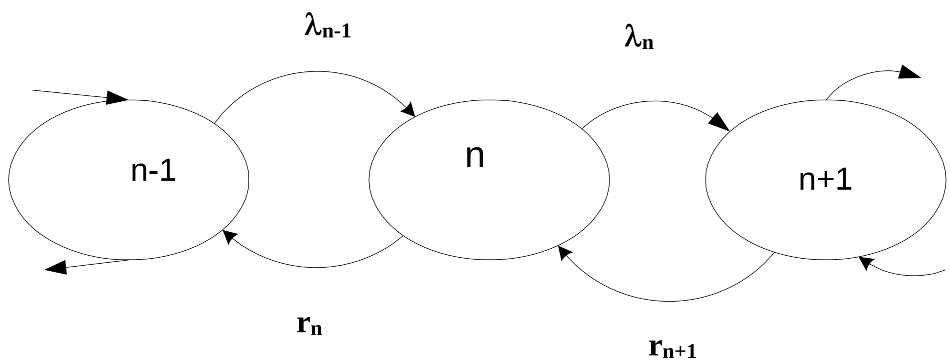


Figure 6. Transition -Rate Flow Diagram for a B&D Process

Figure 6 clearly shows that the flow rate into the state n is given by $\lambda_{n-1}P_{n-1}(t) + r_{n+1}P_{n+1}(t)$ and the rate of transition out of state n is given by $\lambda_n P_n(t) + r_n P_n(t)$. Since $\frac{dP_n(t)}{dt}$ by definition is the rate of change of P_n at time t, then this derivative must equal to

$$\frac{dP_n(t)}{dt} = [\lambda_{n-1}P_{n-1}(t) + r_{n+1}P_{n+1}(t)] - [\lambda_n P_n(t) + r_n P_n(t)] \qquad (82)$$

where flow rate into n is positive and flow rate out of n is taken as negative, as it should. But, the dde in (82) is identical to the dde given in (80d) once $P_n(t)$ is factored out of the last term on the RHS. Further if we just concentrate on the node n = 0, where there can be no death (because of zero no. of failed units) then the flow rate out of n = 0 must equal to $\lambda_0 P_0(t)$ and the transition rate into state n = 0 must equal to $r_1 P_1(t)$ because the population size (or queue length) cannot equal to –1. Hence, the differential equation for node 0 must be $P_0'(t) = \mu_1 P_1(t) - \lambda_0 P_0(t)$, which is identical to (80e). Before we consider different special cases of B&D processes, we will 1st discuss the steady-state long term solution to the system of equations in (80 d & e). By steady-state (or equilibrium) solution we mean as $t \to \infty$, what proportion of the times the system spends in state n = 0, 1, 2 3, 4, ….? Let $\pi_n(t) = \lim_{t \to \infty} P_n(t)$.

Then there are two possibilities: (1) $\pi_n(t)$ will continue to be a function of time in which case the system will never reach equilibrium (or steady-state) and the corresponding limit will be very difficult to compute and may not exist. (2) $\pi_n(t)$ will approach a constant π_n as $t \to \infty$, in which case the $\lim_{t\to\infty} P_n(t)$ will be independent of time, and as a result $\frac{d\pi_n(t)}{dt} = \frac{d\pi_n}{dt} = \frac{d(\text{constant})}{dt} = 0$ in the limit because π_n as $t \to \infty$ will be a constant. Assuming that the system of ddes eventually reaches equilibrium, then as $t \to \infty$, equation (80e) must become: $0 = -\lambda_0 \pi_0 + r_1 \pi_1 \rightarrow \pi_1 = \frac{\lambda_0}{\mu_1}\pi_0$; Equation (80d) for n = 1 now yields: $0 = -(\lambda_1 + r_1)\pi_1 + \lambda_0 \pi_0 + r_2 \pi_2 \rightarrow 0 = -(\lambda_1 + r_1)\frac{\lambda_0}{r_1}\pi_0 + \lambda_0 \pi_0 + r_2 \pi_2 \rightarrow 0 = -(\lambda_1)\frac{\lambda_0}{r_1}\pi_0 + r_2 \pi_2 \rightarrow \pi_2 = \frac{\lambda_0\lambda_1}{r_1 r_2}\pi_0 = \frac{\lambda_1}{r_2}\pi_1$; putting n = 2 in (80d) and letting $t \to \infty$ gives: $0 = -(\lambda_2 + r_2)\pi_2 + \lambda_1 \pi_1 + r_3 \pi_3 \rightarrow 0 = -(\lambda_2 + r_2)\frac{\lambda_1}{r_2}\pi_1 + \lambda_1 \pi_1 + r_3 \pi_3 \rightarrow$ $0 = -(\lambda_2)\frac{\lambda_1}{r_2}\pi_1 + r_3 \pi_3 \rightarrow \pi_3 = \frac{\lambda_1\lambda_2}{r_2 r_3}\pi_1 = \frac{\lambda_2}{r_3}\pi_2 = \frac{\lambda_0\lambda_1\lambda_2}{r_1 r_2 r_3}\pi_0$; similarly, by putting n = 3 in (80d) and letting $t \to \infty$, one will obtain $\pi_4 = \frac{\lambda_3}{r_4}\pi_3 = \frac{\lambda_0\lambda_1\lambda_2\lambda_3}{r_1 r_2 r_3 r_4}\pi_0$, etc. Thus the most general steady-state solution for a B&D process is given by

$$\pi_n = \frac{\lambda_{n-1}}{r_n}\pi_{n-1} = \frac{\lambda_0\lambda_1\lambda_2\lambda_3 \ldots \lambda_{n-1}}{r_1 r_2 r_3 r_4 \ldots r_n}\pi_0 = \pi_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{r_i} \quad , \quad n = 1, 2, 3, \ldots \tag{83a}$$

In order to solve for π_0, we must satisfy the stability constraint $\sum_{n=0}^{\infty} \pi_n \equiv 1$. Applying (83a) to this constraint we obtain:

$$\pi_0 + \sum_{n=1}^{\infty} \pi_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{r_i} = \pi_0 \left(1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{\lambda_{i-1}}{r_i}\right) = 1 \rightarrow$$

$$\pi_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{\lambda_{i-1}}{r_i}} = \frac{1}{1 + \frac{\lambda_0}{r_1} + \frac{\lambda_0\lambda_1}{r_1 r_2} + \frac{\lambda_0\lambda_1\lambda_2}{r_1 r_2 r_3} + \ldots} \tag{83b}$$

In equation (83a) if the sum $\sum_{n=1}^{\infty}\prod_{i=1}^{n}(\frac{\lambda_{i-1}}{r_i})$ is not a finite quantity and does not converge, then $\pi_0 = 1/\infty$ $= 0$ so that the system will never become empty and will grow unboundedly and will become unstable. Then for the system to be stable there must exist an integer k for which $\frac{\lambda_n}{r_{n+1}} < 1$ for all $n \geq k$, or else the number of individuals in the system will approach infinity. Therefore, a strong sufficient condition for a B&D process to be ergodic is that $\frac{\lambda_n}{r_{n+1}} < 1$ for all n = 1, 2, 3, ….. If $\frac{\lambda_n}{r_{n+1}} = 1$ for all $n \geq k$, then all states become null recurrent; $\frac{\lambda_n}{r_{n+1}} > 1$ all states become transient. For the case of identical exponential arrival and departure times, i.e., when all $\lambda_n = \lambda$ (a constant arrival or failure rate) for all n = 0, 1, 2, 3, …, and $r_n = \lambda_r$ (a constant service or repair rate) for all n = 1, 2, 3, …, equations (80a &b) reduce to $\pi_n = (\frac{\lambda}{\lambda_r})^n \pi_0 = \rho^n \pi_0$, n = 0, 1, 2, 3, 4 , and $\sum_{n=0}^{\infty}\pi_n = \sum_{n=0}^{\infty}\rho^n \pi_0 = \pi_0 \sum_{n=0}^{\infty}\rho^n = \frac{\pi_0}{1-\rho} = 1$ iff $0 \leq \rho < 1 \rightarrow \pi_0 = 1-\rho$. Therefore, for this special case the pmf of number of members in the system (if $0 \leq \rho < 1$) is given by

$$\pi_n = (1-\rho)\,\rho^n,\ \ n = 0, 1, 2, 3 \ldots. \qquad \textbf{(83c)}$$

which is a geometric process at a traffic intensity $0 < \rho = \lambda/\lambda_r < 1$. The Pr of finding the system empty in the long run is given by $\pi_0 = 1-\rho$. The average number of customers (or failed units) in the system in the long run is given by the geometric mean $E(N) = \sum_{n=0}^{\infty} n(1-\rho)\rho^n = \frac{\rho}{1-\rho}$ and the variance of the geometric pmf in (83c) is given by $V(N) = E(N^2) - (\frac{\rho}{1-\rho})^2 = \sum_{n=0}^{\infty} n^2(1-\rho)\rho^n - (\frac{\rho}{1-\rho})^2 = \frac{\rho}{(1-\rho)^2}$.

The average total time (both queueing and service times) spent in the system can easily be obtained from J. D. C. Little's Result (1961): $E(N) = \lambda\, E(T_{Sys})$, which so simply states that the average number of customers in a any queueing system is directly proportional with the average arrival rate times the average amount of time spent in the system per customer. Applying the Little's result to the most fundamental system M/M/1, we obtain $\frac{\rho}{1-\rho} = \lambda\, E(T_{Sys}) \rightarrow E(T_{Sys}) = \frac{\rho}{\lambda(1-\rho)} = \frac{\lambda/\lambda_r}{\lambda(1-\lambda/\lambda_r)} =$

$$\frac{1}{\lambda_r - \lambda} = \frac{1/\lambda_r}{(1-\lambda/\lambda_r)} = \frac{\textbf{Average service time}}{\pi_0}.$$ This last relationship states that the average system time is inversely proportional to proportion of the times that the system is empty but directly proportional to average service time. Again the reader must observe that since $E(T_{Sys}) = \frac{1}{\lambda_r - \lambda} > 0$, this again shows that the mean repair (or service rate) for any system has to exceed the mean arrival rate for system to remain ergodic, or else it will become unstable.

Now that we have laid down the foundation for simple queueing systems, specific cases follow.

Repairable Systems

Chapter 5 of Ebeling dealt only with system RE W/O repair and maintenance, and therefore, system availability at time t, A(t), was simply the same as system RE at time t. Examples of irreparable components are light bulbs, resistors, batteries and computer chips, while complex systems such as airplanes, cars and air conditioning systems have many repairable components. If a system is repairable (either on-line or off-line, where by on-line we mean operations will not be interrupted), then there are three important performance criteria: MTBF (Mean Time Between Failures), and steady-state intrinsic availability A_I, and dependability $D(t) = A_I \times R(t)$. By renewal we mean that a system fails but upon failure the failed component is either replaced with a brand new unit or is repaired to its original condition. This is called a renewal process. A renewal process is the generalization of a Poisson process where the interarrival (or intervening) times between two successive events (failures) can have any pdf instead of just the exponential.

If the failed component is immediately replaced with a new one (or if its Time-to-Repair, TTR, is negligible), then the long-term availability of the system is almost 100%, and its time-dependent availability, A(t), is equal to $R_{Sys}(t)$. Otherwise, if the TTR has a specified distribution, such as exponential with repair rate λ_r or lognormal, then the steady-state (i.e., as $t \to \infty$) intrinsic availability was shown to equal the expression given below.

$$\mathbf{A_I = \frac{MTBF}{MTBF + MTTR}} \qquad \textbf{(84a)}$$

If the system failure rate is a constant λ and its restoration rate is also a constant λ_r, then (84a) reduces to

$$A_I = \frac{1/\lambda}{1/\lambda + 1/\lambda_r} = \frac{\lambda_r}{\lambda_r + \lambda} \; . \qquad \textbf{(84b)}$$

Further, by on-line restoration we mean that component failure does not interrupt system functioning. If on-line restoration is not possible, then the entire system has to fail first before system restoration is performed off-line at the system restoration rate λ_r. Note that in the case of off-line restoration λ_r = r represents the restoration rate of the entire system (not individual components).

The Two-State Markov Process

For the sake of specificity, consider the network system example given on pages 109-110 of these notes but assume that now time is continuous (i.e., the system is monitored on a continual basis). The average failure rate (downtime rate) of the network is λ = 0.004 per hour and a repair rate λ_r = 0.08/hour. Thus the MTBF = 1/ λ = 250 hours and MTBR = 1/λ_r = 12.5 hours. The C-K dde for node 0 is

$$P_{00}(t + \Delta t) = P_{00}(t)\times[1 - \lambda\, \Delta t] + P_{01}(t)\,(\lambda_r \Delta t)\,[1 - \lambda\, \Delta t]$$

After transposing $P_{00}(t)$ to the LHS, dividing by Δt, and then letting $\Delta t \to$ o, we obtain

$$\frac{dP_{00}(t)}{dt} = -\lambda\, P_{00}(t) + \lambda_r\, P_{01}(t) \to P'_{00}(t) + \lambda\, P_{00}(t) = \lambda_r\, P_{01}(t) = \lambda_r\, [1 - P_{00}(t)] \to$$

$$P'_{00}(t) + (\lambda + \lambda_r)\, P_{00}(t) = \lambda_r \to P'_{00}(t)\, e^{(\lambda+\lambda_r)t} + (\lambda + \lambda_r)\, e^{(\lambda+\lambda_r)t}\, P_{00}(t) =$$

$$\lambda_r\, e^{(\lambda+\lambda_r)t} \to \frac{d}{dt}[P_{00}(t)e^{(\lambda+\lambda_r)t}] = \mu\, e^{(\lambda+\lambda_r)t} \to d[P_{00}(t)e^{(\lambda+\lambda_r)t}] = \mu\, e^{(\lambda+\lambda_r)t}dt \to$$

$$\int d[P_{00}e^{(\lambda+\lambda_r)t}] = \int \mu e^{(\lambda+\lambda_r)t}dt + C \to P_{00}(t)e^{(\lambda+\lambda_r)t} = \frac{\lambda_r}{\lambda+\lambda_r}\, e^{(\lambda+\lambda_r)t} + C \to$$

$$P_{00}(t) = \frac{\lambda_r}{\lambda+\lambda_r} + C\; e^{-(\lambda+\lambda_r)t} \qquad \textbf{(85a)}$$

Applying the boundary condition $P_{00}(t = 0) = 1$ to the solution in (85a) results in 1 =

$\frac{\lambda_r}{\lambda+\lambda_r} + C \to C = 1 - \frac{\lambda_r}{\lambda+\lambda_r} = \frac{\lambda}{\lambda+\lambda_r}$ and the final solution

$$P_{00}(t) = \frac{\lambda_r}{\lambda+\lambda_r} + \frac{\lambda}{\lambda+\lambda_r} e^{-(\lambda+\lambda_r)t}, \text{ and } P_{01}(t) = \frac{\lambda}{\lambda+\lambda_r} - \frac{\lambda}{\lambda+\lambda_r} e^{-(\lambda+\lambda_r)t} \quad \textbf{(85b)}$$

If the network starts in state 1 (down), then

$$P_{10}(t+\Delta t) = P_{10}(t)\times[1-\lambda\,\Delta t] + P_{11}(t)(\lambda_r\,\Delta t) \rightarrow P'_{10}(t) + \lambda P_{10}(t) = \lambda_r P_{11}(t) = \lambda_r[1-P_{10}(t)] \rightarrow$$

$$P'_{10}(t) + (\lambda+\lambda_r)P_{10}(t) = \lambda_r \rightarrow P_{10}(t) = \frac{\lambda_r}{\lambda+\lambda_r} + C\, e^{-(\lambda+\lambda_r)t};$$ **applying the boundary condition**

$P_{10}(t=0) = 0$, **we obtain** $0 = \frac{\lambda_r}{\lambda+\lambda_r} + C \rightarrow C = -\frac{\lambda_r}{\lambda+\lambda_r} \rightarrow$

$$P_{10}(t) = \frac{\lambda_r}{\lambda+\lambda_r} - \frac{\lambda_r}{\lambda+\lambda_r} e^{-(\lambda+\lambda_r)t}, \text{ and } P_{11}(t) = \frac{\lambda}{\lambda+\lambda_r} + \frac{\lambda_r}{\lambda+\lambda_r} e^{-(\lambda+\lambda_r)t} \quad \textbf{(85c)}$$

Thus, $\pi_0 = \lim_{t\to\infty} P_{00}(t) = \lim_{t\to\infty} P_{10}(t) = \frac{\lambda_r}{\lambda+\lambda_r}$, and $\pi_1 = \lim_{t\to\infty} P_{01}(t) = \lim_{t\to\infty} P_{11}(t) = \frac{\lambda}{\lambda+\lambda_r}$. The reader must observe that $P_{00}(t)$ is equal to system reliability at time t given that the network is initially in an operational state, and $P_{10}(t)$ is the reliability function at time t given that the system starts in the failed state.

Substituting $\lambda = 0.004$ per hour and a repair rate $\lambda_r = 0.08$/hour for our specific example, we obtain $\pi_0 = \frac{\lambda_r}{\lambda+\lambda_r} = \frac{0.08}{0.004+0.08} = 80/84 = 20/21$ and $\pi_1 = 1/21$. Thus, on the average the network is down one hour every 21 hours, and the long-term availability is given by $A_I = \pi_0 = \frac{\lambda_r}{\lambda+\lambda_r} = 20/21$. The Pr that the network will be under maintenance in 3 days given that presently is in state 0 is given by

$$P_{01}(t = 72 \text{ hours}) = \frac{\lambda}{\lambda+\lambda_r} - \frac{\lambda}{\lambda+\lambda_r} e^{-(\lambda+\lambda_r)t} = \frac{0.004}{0.004+0.08} - \frac{0.004}{0.004+0.08} e^{-(0.084)72} = \frac{1}{21} -$$

$$\frac{1}{21} e^{-6.048} = 0.0475065 \rightarrow P_{00}(t=72) = 1 - 0.04751 = 0.9524935.$$ **Similarly,** $P_{10}(t = 72) =$

$$\frac{0.08}{0.004+0.08} - \frac{0.08}{0.084} e^{-6.048} = 0.950131.$$

Note that our system reliability with maintenance at 72 hours, given that initially the system is in the up state, is equal to $R_{00}(72) = 0.9524935$. Without maintenance the same system reliability would be equal to $R(72) = \mathbf{e}^{-\lambda t} = \mathbf{e}^{-0.004(72)} = 0.7497616$. For this reason when a system includes maintenance, the functions $P_{00}(t)$ and $P_{10}(t)$ yield much higher system reliability. The expected length of time the network would be under repair in 72 hours, given that initially it is in state 0, is given by

$$\text{MTTR}(01) = \int_0^{72} \left[\frac{\lambda}{\lambda+\lambda_r} - \frac{\lambda}{\lambda+\lambda_r}\mathbf{e}^{-(\lambda+\lambda_r)t}\right]\mathbf{dt} = \frac{72\lambda}{\lambda+\lambda_r} + \frac{\lambda}{(\lambda+\lambda_r)^2}\left[e^{-6.048} - 1\right] = 2.86302 \text{ hours,}$$

while the value of $\text{MTTR}_{11} = 14.7397$ hours.

Deriving the Transition Rate Matrix for a 2-State Markov Chain

In the context of the above network example, let us assume that our time zero is always an epoch of time when the system is in state 0 (or operating reliably), i.e., in this section we assume that the $P_0(t = 0) = 1$. Then our C-K ddes become

$$\begin{cases} \mathbf{P}_0'(\mathbf{t}) = -\lambda \mathbf{P}_0(\mathbf{t}) + \lambda_r \mathbf{P}_1(\mathbf{t}) \\ \mathbf{P}_1'(\mathbf{t}) = \lambda \mathbf{P}_0(\mathbf{t}) - \lambda_r \mathbf{P}_1(\mathbf{t}) \end{cases}.$$

Note that the sum of these ddes must add to zero because $\frac{\mathbf{d}}{\mathbf{dt}}[\mathbf{P}_0(\mathbf{t}) + \mathbf{P}_1(\mathbf{t})] = \frac{\mathbf{d}}{\mathbf{dt}}[\mathbf{1.0000}] \equiv 0$. We may now write this last system of the two dde in matrix form as follows:

$$\begin{bmatrix} \mathbf{P}_0'(\mathbf{t}) \\ \mathbf{P}_1'(\mathbf{t}) \end{bmatrix} = \begin{bmatrix} -\lambda & \lambda_r \\ \lambda & -\lambda_r \end{bmatrix} \times \begin{bmatrix} \mathbf{P}_0(\mathbf{t}) \\ \mathbf{P}_1(\mathbf{t}) \end{bmatrix} \rightarrow \mathbf{P}'(\mathbf{t}) = \mathbf{B} \times \mathbf{P}(\mathbf{t}) \qquad \textbf{(86)}$$

where $B = \begin{bmatrix} -\lambda & \lambda_r \\ \lambda & -\lambda_r \end{bmatrix}$ is called the transition rate matrix (TRM) and different authors use different notations for the TRM. The definition of the 2×1 vectors $\mathbf{P}'(\mathbf{t})$ and $P(t)$ should be self-explanatory in equation (86). Solving the system of ddes in (86) as before we obtain

$$A(t) = P_0(t) = \frac{\lambda_r}{\lambda+\lambda_r} + \frac{\lambda}{\lambda+\lambda_r}\mathbf{e}^{-(\lambda+\lambda_r)t}, \text{ and } Q(t) = U(t) = P_1(t) = \frac{\lambda}{\lambda+\lambda_r} - \frac{\lambda}{\lambda+\lambda_r}\mathbf{e}^{-(\lambda+\lambda_r)t},$$

where U(t) is called the system unavailability at time t.

We now solve the dde (86) using matrix manipulations. We first observe that the infinite Maclaurin series of matrices $\mathbf{e^{Bt}} = \mathbf{I + Bt} + \frac{\mathbf{B^2}}{\mathbf{2!}}\mathbf{t^2} + \frac{\mathbf{B^3}}{\mathbf{3!}}\mathbf{t^3} + \ldots = \sum_{\mathbf{n=0}}^{\infty} \frac{\mathbf{(Bt)^n}}{\mathbf{n!}}$, called the state transition matrix, actually satisfy the matrix differential equations in (86), if we adopt the initial condition $P(t = 0) = \mathbf{e^0} = I = \frac{\mathbf{(Bt)^0}}{\mathbf{0!}}$. It can also be shown that almost any square matrix can be diagonalized thru $B = \mathbf{M}\Lambda\mathbf{M}^{-1}$, where Λ is a diagonal matrix consisting of eigenvalues of B and the matrix **M** may be an oblique (i.e., nonorthogonal) matrix whose columns are the matrix B's eigenvectors. Thus, the general solution is $P(t) = \mathbf{M}\mathbf{e}^{\Lambda t}\mathbf{M}^{-1}$, where $\mathbf{e}^{\Lambda t} = \begin{bmatrix} \mathbf{e}^{\xi_1 \mathbf{t}} & \mathbf{0} \\ \mathbf{0} & \mathbf{e}^{\xi_2 \mathbf{t}} \end{bmatrix}$ and ξ_i (i =1, 2) are the eigenvalues of B, i.e., $\Lambda = \begin{bmatrix} \xi_1 & \mathbf{0} \\ \mathbf{0} & \xi_2 \end{bmatrix}$. Using matrix manipulations, our transient solution at t = 72 hours is given by $P(72) = \mathbf{M}\ \mathbf{e}^{\Lambda t}\mathbf{M}^{-1} =$

$$\begin{bmatrix} \mathbf{0.99875234} & \mathbf{-0.70710678} \\ \mathbf{0.04993762} & \mathbf{0.70710678} \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{0.0023626} \end{bmatrix} \times \begin{bmatrix} \mathbf{0.95357107} & \mathbf{0.953571} \\ \mathbf{-0.0673435} & \mathbf{1.346870} \end{bmatrix}$$

$= \begin{bmatrix} \mathbf{0.95249346} & \mathbf{0.950130874} \\ \mathbf{0.04750654} & \mathbf{0.049869126} \end{bmatrix}$. Note that if we formally write the matrix ddes $\mathbf{P'(t)} = B \times P(t)$ in the form $dP/P(t) = B\,dt$, then we would recognize that the solution has to be of the form $\ln[P(t)] = B\,t + \ln(c)$. However, these operations are disallowed because we are dealing with matrix differential equations.

Example 12. To further illustrate the applications of the above procedures to a chain with more than two states, consider the Markovian system in a small airport which handles a fleet of 4 aircrafts (as I made up this example I had to limit the fleet size to 4 just for illustrative purposes). The fleet is maintained by a crew of 2 servicemen on a FCFS (first-come, first-serve) priority basis. The mean time between downtimes for each aircraft is MTBF = 30 hours and the mean time to repair (MTTR) for a single aircraft worked on both servicemen is MTTR = 2.5 hours. The following TRD describes the system behavior, where X(t) represents the number of aircrafts that are down (i.e., either being repaired or waiting in queue).

Figure 7 clearly shows that $\mathbf{P}_0'(\mathbf{t})$ = – (4/30) P_0(t) + (2/5) P_1(t); $\mathbf{P}_1'(\mathbf{t})$ = – (3/30+ 2/5) P_1(t) + 4/30 P_0(t) + 2/5 P_2(t); $\mathbf{P}_2'(\mathbf{t})$ = – (2/30+ 2/5) P_2(t) + 3/30 P_1(t) + 2/5 P_3(t); $\mathbf{P}_3'(\mathbf{t})$ = – (1/30+ 2/5) P_3(t) + 2/30 P_2(t) + 2/5 P_4(t); $\mathbf{P}_4'(\mathbf{t})$ = – (2/5) P_4(t) + 1/30 P_3(t); The transient (or time-dependent) solution for this 5-state ddes is very difficult to obtain. I will take a stab at it for just a short while, but henceforth we will concentrate more on the steady-state solution π_n (n = 0, 1, 2, 3, 4), i.e., as t →

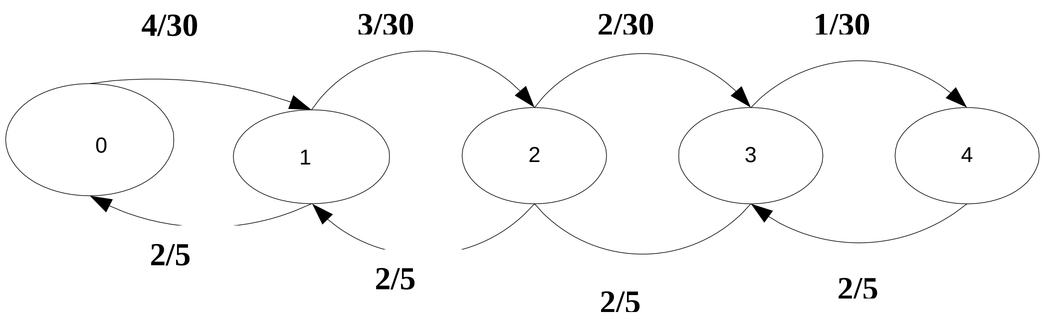


Figure 7. The TRD for the fleet of 4 aircrafts

∞. Before starting the discussion about obtaining the transient solution, the reader should observe that the average death (or repair) rate in figure 7 is clearly μ = 2/5 per aircraft, but the average arrival rate $\bar{\lambda}$ will be determined later.

The Transient (or Time-Dependent) Solution

As stated earlier the general solution as a function of time is given by P(t) = $\mathbf{M}e^{\Lambda t}\mathbf{M}^{-1}$, where

$$\Lambda = \begin{bmatrix} \xi_1 & 0 & 0 & 0 & 0 \\ 0 & \xi_2 & 0 & 0 & 0 \\ 0 & 0 & \xi_3 & 0 & 0 \\ 0 & 0 & 0 & \xi_4 & 0 \\ 0 & 0 & 0 & 0 & \xi_5 \end{bmatrix}$$

, and ξ_i (i =1, 2, 3, 4, 5) are the eigenvalues of

the transition rate matrix $\mathbf{B} = \begin{bmatrix} -4/30 & 2/5 & 0 & 0 & 0 \\ 4/30 & -1/2 & 2/5 & 0 & 0 \\ 0 & 3/30 & -14/30 & 2/5 & 0 \\ 0 & 0 & 2/30 & -13/30 & 2/5 \\ 0 & 0 & 0 & 1/30 & -2/5 \end{bmatrix}$.

Note that the way I have defined the TRM (transition rate matrix), it does satisfy $\mathbf{P'(t)} = \mathrm{B}\times\mathrm{P(t)}$, where

$\mathrm{P(t)} = \begin{bmatrix} P_0(t) \\ P_1(t) \\ P_2(t) \\ P_3(t) \\ P_4(t) \end{bmatrix}$; further, the columns of the TRM must always add to exactly zero, and it should be read

from column to rows (unlike the transition pr matrix P). The eigenvalues of the matrix B are (from Matlab) $\xi_0 = 0.00$, $\xi_1 = -0.7700597$, $\xi_2 = -0.2230742$, $\xi_3 = -0.3827541$, $\xi_5 = -0.5574453$, and the eigenvector matrix pertaining to these 5 eigenvalues is **M** =

$\begin{bmatrix} 0.94565 & -0.49322 & 0.86082 & -0.77477 & 0.66595 \\ 0.31522 & 0.78511 & -0.19313 & 0.48311 & -0.70608 \\ 0.07881 & -0.36566 & -0.42065 & 0.39986 & -0.12058 \\ 0.01314 & 0.08107 & -0.20788 & -0.03689 & 0.20389 \\ 0.00110 & -0.00730 & -0.03917 & -0.07131 & -0.04317 \end{bmatrix}$. Thus the transient (or time-dependent) solution is given by

$$\mathbf{P(t) = M\, e^{\Lambda t} M^{-1} = M} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & e^{-0.77006t} & 0 & 0 & 0 \\ 0 & 0 & e^{-0.22308t} & 0 & 0 \\ 0 & 0 & 0 & e^{-0.38276t} & 0 \\ 0 & 0 & 0 & 0 & e^{-0.55745t} \end{bmatrix} \mathbf{M^{-1}}.$$

Note that you may not reverse the order of the $\mathbf{M}$ matrix in the above solution, i.e., $P(t) \neq \mathbf{M}^{-1} \mathbf{e}^{\Lambda t} \mathbf{M}$, where this last expression is completely erroneous. The time-dependent solution for 20 hours is given by

$$\mathbf{P}(20) = \begin{bmatrix} 0.6997 & 0.6976 & 0.6915 & 0.6785 & 0.6560 \\ 0.2325 & 0.2330 & 0.2345 & 0.2372 & 0.2412 \\ 0.0576 & 0.0586 & 0.0616 & 0.0679 & 0.0789 \\ 0.0094 & 0.0099 & 0.0113 & 0.0146 & 0.0208 \\ 0.0008 & 0.0008 & 0.0011 & 0.0017 & 0.0031 \end{bmatrix}.$$

The 1st row of the above matrix gives the pr of finding the system in state 0 at t = 20; the 2nd row of the matrix gives the pr of finding the system in state 1; the 3rd row of the above matrix gives the pr of finding the system in state 2 at time t = 20, etc.

I hope the reader now appreciates the complexity of obtaining the transient solutions to a set of ddes. Thus, we obtain the steady-state (i.e., as $t \to \infty$) for the fleet of 4 aircrafts. We have to solve the system of $B \times \pi = 0$ imposing the constraint $\sum_{n=0}^{4} \pi_n \equiv 1$. $\to$

$$\begin{bmatrix} -4/30 & 2/5 & 0 & 0 & 0 \\ 4/30 & -1/2 & 2/5 & 0 & 0 \\ 0 & 3/30 & -14/30 & 2/5 & 0 \\ 0 & 0 & 2/30 & -13/30 & 2/5 \\ 0 & 0 & 0 & 1/30 & -2/5 \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \to$$

$$\begin{cases} -\pi_0 + 3\pi_1 = 0 \\ 4\pi_0 - 15\pi_1 + 12\pi_2 = 0 \\ 3\pi_1 - 14\pi_2 + 12\pi_3 = 0 \\ 2\pi_2 - 13\pi_3 + 12\pi_4 = 0 \\ \pi_0 + \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 \end{cases} \quad \to \quad \pi = \begin{bmatrix} 0.69846402586904 \\ 0.23282134195635 \\ 0.05820533548909 \\ 0.00970088924818 \\ 0.00080840743735 \end{bmatrix}$$

The average number of aircrafts in the system (being down) in the long-run is given by E(N) = $\sum_{n=0}^{4} n\pi_n$ = 0.38157 aircrafts. The average arrival rate for repair is given by $\bar{\lambda} = \sum_{n=0}^{4} \pi_n \lambda_n$ = $\frac{4}{30}\times\pi_0 + \frac{3}{30}\times\pi_1 + \frac{2}{30}\times\pi_2 + \frac{1}{30}\times\pi_3 + 0\times\pi_4$ = 0.1206144 units/hour $\rightarrow \bar{N}$ = E(N) = $\bar{\lambda}\times\bar{T}_{Sys}$ (Little's Result) $\rightarrow \bar{T}_{Sys} = \bar{N}/\bar{\lambda}$ = 0.38157/0.1206144 = 3.1635389 hours. Since in general $\bar{T}_{Sys} = \overline{W}$ + MTTR = Average Waiting Time + Average Service Time, then in follows that the average queueing time per aircraft is given by $\overline{W} = \bar{T}_{Sys}$ – MTTR = 3.1635389 – 2.5 = 0.6635389 hours.

Exercise23. As a generalization of the above fleet of 4 aircrafts problem, consider an air force base where aircrafts arrive at a rate of λ (say λ = 1.20 per hour) and the Pr that a landed aircraft needs an un-routine maintenance is p (say p = 1/3). The base maintenance shop can repair an airplane at the following rates λ_r : r_1 = 0.35 per hour, or r_2 = 0.50 per hour. Draw the TRD and model this situation as a B&D process. If the shop manager adopts = 0.35 per hour, would he be considered a competent supervisor? Answers: $\pi_n = (\frac{\lambda p}{\lambda_r})^n \pi_0$; $\sum_{n=0}^{\infty}\pi_n \equiv 1 \rightarrow \pi_0 = 1 - \frac{\lambda p}{\lambda_r} = 1 - \rho \rightarrow$ Traffic Intensity in the repair facility = $\rho = \frac{\lambda p}{\lambda_r} = \frac{1.20\times 1/3}{\lambda_r} = \frac{0.40}{\lambda_r} \rightarrow$ The supervisor is making the wrong decision to set μ = 0.35 because then ρ = 1.1429 > 1 implying that the queue length, Q_n, will grow unlimitedly and will lead to a very unstable situation. Therefore, the manager must adopt λ_r = 0.5 per hour in which case ρ = 0.80, and the pr of system being empty will equal to π_0 = 0.20. When λ_r = 0.50, the steady-state solution is $\pi_n = (0.80^n)\,\pi_0$, and the average number of aircrafts in the system is given by E(N) = $\bar{N}$ =

$$\sum_{n=0}^{\infty} n\pi_n = .20\sum_{n=1}^{\infty} n(0.8)^n = 00.16\sum_{n=1}^{\infty} n(0.8)^{n-1} = 0.16\sum_{n=1}^{\infty} n(\rho)^{n-1} = 0.16\frac{d}{d\rho}\sum_{n=1}^{\infty}\rho^n =$$

$0.16\ \frac{d}{d\rho}(\frac{\rho}{1-\rho}) = 0.16/(1-\rho)^2 = 4$ aircrafts. Using Little's result, we obtain $\bar{N} = \bar{\lambda} \times \bar{T}_{Sys} \rightarrow 4 = 0.4\ \bar{T}_{Sys} \rightarrow \bar{T}_{Sys} = 10.00$ hours; since the mean repair time is MTTR = $1/\lambda_r = 2$ hours, then the average queueing time is given by $\overline{W} = 8$ hours. If we consider the queue length itself, then the average queue length is given by $\bar{N}_q = \bar{\lambda} \times \overline{W} = 0.40 \times 8 = 3.2$ aircrafts in the repair facility.

Reliability Analysis of a Repairable Two-Component System

Such a system can be either a series or a pure parallel system. We work out the Markov analysis in general and will obtain the availability functions for the two possibilities. The TRD will be given below. I have borrowed this example from the text by R. Ramakumar, "Engineering Reliability", Prentice Hall, ISBN: 0-13-276759-7, pp. 264-266, but my analysis does not follow the same path as the author as I found the presentation in that book not quite sufficiently specific. We must state up front whether there is a single repairman (single server) or two repairmen because when both units are down (or are in the failed state DD), we have to decide if the repair (or death) rate is either λ_r, or $r_1 + r_2$, where the former pertains to one serviceman while the latter pertains to two service facility. I will work out the details of the former case with a single server first, then followed by the latter, assuming that when both units are down (state DD), unit 1 is repaired first. My state TRD for the 1-server case is given Figure 8, where there are 4 possible states of the system: State 1 = both units are up = UU, State 2 = unit 1 is down but unit 2 is up = DU, State 3 = unit 1 is up but unit 2 is down = UD, State 4 = both units are down = DD. As I have illustrated above obtaining transient solutions to such systems is extremely difficult, and even if we obtain them, then we would want to let $t \rightarrow \infty$ to obtain the steady –state (or equilibrium) solution. Therefore, we may as well assume that $P_i'(t) = 0$, for all i =1, 2, 3, 4 and equate the flow rate into a state to the flow rate out of the same state, as given below. The transition rate out of state 1 is $\lambda_1 + \lambda_2$ and the corresponding transition rate into state 1 is equal to either r_1 or r_2 (not both because during Δt it is impossible for two repairs to occur). Weighing each of these transition rates by the corresponding π_i, and equating the corresponding flow rates we obtain $(\lambda_1 + \lambda_2)\ \pi_1 = r_1\ \pi_2 + r_2\ \pi_3$. Similarly, equating the flow rates in and out of state 2 leads to $(\lambda_2 + r_1)\ \pi_2 = \lambda_1\ \pi_1 + r_2\ \pi_4$. For the state 4 we have: $(r_1 + r_2)\ \pi_4 = \lambda_1\ \pi_3 + \lambda_2\ \pi_2$. These lead to the following system of 4 equations with 4 unknowns.

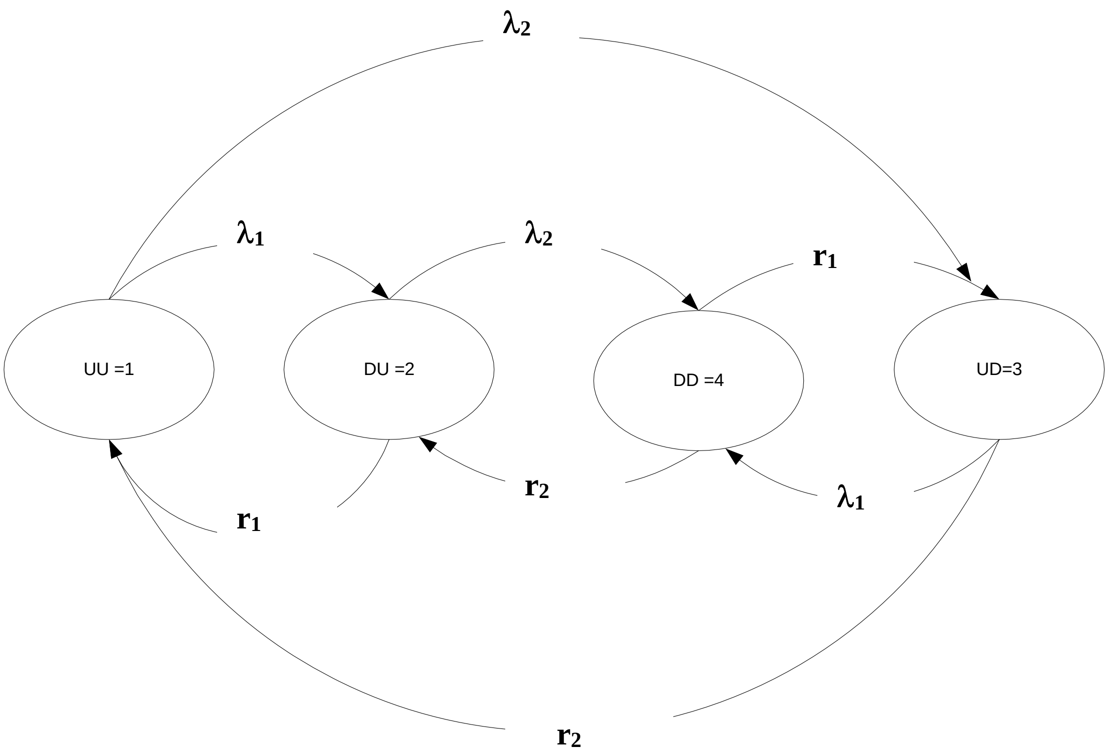


Figure 8. State Transition Rate diagram for a 2-unit On-line Repairable System

$$\begin{cases} -(\lambda_1+\lambda_2)\pi_1 + r_1\pi_2 + r_2\pi_3 = 0 \\ \lambda_1\pi_1 - (\lambda_2 + r_1)\pi_2 + r_2\pi_4 = 0 \\ \lambda_2\pi_2 + \lambda_1\pi_3 - (r_1 + r_2)\pi_4 = 0 \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1 \end{cases} \qquad (87)$$

The above system of equations can be solved in general using Cramer's Rule, which is somewhat painful. So, I will make up some transition rates and provide the corresponding solution. Suppose that λ_1 = 0.01 per hour, λ_2 = 0.015 per hour, r_1 = 0.50 per hour, and r_2 = 0.40 per hour. Substituting these transition rates into equation (87) and solving the resulting system of equations **yields** $\boldsymbol{\pi = A^{-1} \times b}$**,**

where $A = \begin{bmatrix} -0.025 & 0.5000 & 0.4000 & 0 \\ 0.01 & -0.51500 & 0 & 0.40 \\ 0 & 0.0150 & 0.01 & -0.90 \\ 1.000 & 1.00 & 1.00 & 1.00 \end{bmatrix}$, $b = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$, and

π = [0.944956296 0.018899126 0.0354358611 0.0007087172]′. If the repairable 2-unit system is in series, then the system long-term average availability is given by $A_I = \pi_1 \cong 0.9449563$, but if it is a pure parallel system, then $A = \pi_1 + \pi_2 + \pi_3 = 0.999291283$, and its $U = \pi_4 = 0.000708717$.

Now consider the special case of a 2-unit on-line repairable system, where the two units are identical so that $\lambda_1 = \lambda_2 = \lambda$, and $r_1 = r_2 = \lambda_r$ with 2 servers; let X(t) represent the number of units that are down (or are in the failed state) at time t. Then the state TRD in figure 8 reduces to the one in figure 9. From the TRD in Figure 9 we can easily deduce that $2\lambda\pi_0 = \lambda_r \pi_1 \rightarrow \pi_1 = (2\lambda/\lambda_r)\pi_0$;

$(\lambda + \lambda_r)\pi_1 = 2\lambda\, \pi_0 + 2\lambda_r\, \pi_2 \rightarrow (\lambda + \lambda_r)\dfrac{2\lambda}{\lambda_r}\pi_0 = 2\lambda\, \pi_0 + 2\lambda_r \pi_2 \rightarrow \dfrac{2\lambda^2}{\lambda_r}\pi_0 = 2\lambda_r \pi_2 \rightarrow \pi_2 =$

$(\lambda/\lambda_r)^2\pi_0 \rightarrow \pi_0 + (2\lambda/\lambda_r)\pi_0 + (\lambda/\lambda_r)^2\pi_0 = 1 \rightarrow \pi_0 = \dfrac{1}{1+(2\lambda/\lambda_r)+(\lambda/\lambda_r)^2} =$

$\dfrac{\lambda_r^2}{\lambda_r^2 + 2\lambda\lambda_r + \lambda^2} = \dfrac{\lambda_r^2}{(\lambda+\lambda_r)^2} = [\lambda_r/(\lambda_r+\lambda)]^2$. Once π_0 is computed, the other two steady-state prs can easily be obtained. For example, if $\lambda = 0.015$ and $\lambda_r = 0.50$ per hour, then $\pi_0 = 0.94259590913$, $\pi_1 = 0.05655575455$, and $\pi_2 = 0.00084833632$. Then for a series system the average system availability is $A = \pi_0 = 0.94259590913$, and $U = 0.05740409087$. However, if the system is a pure parallel one, then $A_I = 0.99915166368$ and $U = 0.00084833632 = [\lambda/(\lambda_r+\lambda)]^2$. The special case of the above is a 2-unit series system with only one server, i.e., the transition rate out of state 2 back to 1 is given by $r_{21} = \lambda_r$ (not $2\lambda_r$). Then, $2\lambda\pi_0 = \lambda_r\, \pi_1 \rightarrow \pi_1 = (2\lambda/\lambda_r)\pi_0$; $(\lambda + \lambda_r)\pi_1 = 2\lambda\, \pi_0 + \lambda_r \pi_2$

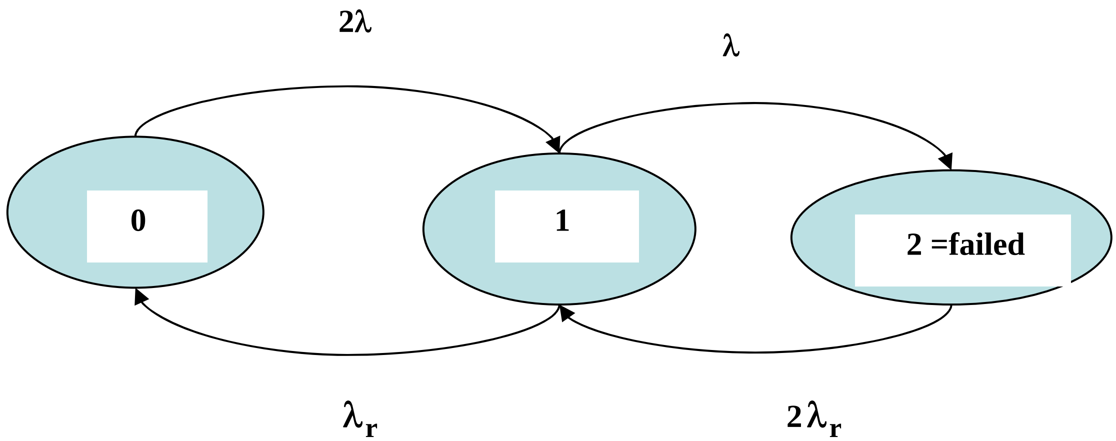


Figure 9. Two-Unit On-Line Repairable System

$\rightarrow (\lambda + \lambda_r)\, \dfrac{2\lambda}{\lambda_r}\pi_0 = 2\lambda\, \pi_0 + \lambda_r\pi_2 \;\rightarrow\; \dfrac{2\lambda^2}{\lambda_r}\pi_0 = \lambda_r\pi_2 \;\rightarrow \pi_2 = (2\lambda^2/\,\lambda_r^2)\pi_0$

$\rightarrow \pi_0 + (2\lambda/\lambda_r)\pi_0 + 2(\lambda/\lambda_r)^2\,\pi_0 = 1 \rightarrow \pi_0 = \dfrac{1}{1+(2\lambda/\lambda_r)+2(\lambda/\lambda_r)^2} = \dfrac{\lambda_r^2}{\lambda_r^2+2\lambda\lambda_r+2\lambda^2}$. Thus, the average availability in the long run for a 2-unit repairable series system is given by $A = \pi_0 = \dfrac{\lambda_r^2}{\lambda_r^2+2\lambda\lambda_r+2\lambda^2}$.

To compute the $MTTF_{Sys}$ for a 2-unit parallel repairable system, we have to assume that the state 2 in figure 9 is absorbing so that we may obtain the fundamental matrix N. Figure 9 shows that the transition Pr $P_{01} = (2\lambda)\,\Delta t$ and thus $P_{00} = 1 - (2\lambda)\,\Delta t$. Similarly, $P_{10} = \lambda_r(\Delta t)$, $P_{12} = \lambda\,(\Delta t)$, and hence $P_{11} = 1 - \lambda_r(\Delta t) - \lambda\,(\Delta t) = 1 - (\lambda_r + \lambda)\Delta t$. If we take Δt equal to 1 unit of time, then our one-step transitional pr matrix for a duration Δt is given by

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 1-2\lambda & 2\lambda & 0 \\ \lambda_r & 1-\lambda_r-\lambda & \lambda \\ 0 & 2\lambda_r & 1-2\lambda_r \end{bmatrix} \end{array}.$$

Making state 2 absorbing, this last matrix reduces to

$$\tilde{\mathbf{P}} = \begin{array}{c} \\ 2 \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} 2 & 0 & 1 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-2\lambda & 2\lambda \\ \lambda & \lambda_r & 1-\lambda-\lambda_r \end{bmatrix} \end{array} = \begin{bmatrix} \mathbf{P}_1 & 0 \\ \mathbf{R} & \mathbf{Q} \end{bmatrix} \rightarrow \mathbf{Q} = \begin{bmatrix} 1-2\lambda & 2\lambda \\ \lambda_r & 1-\lambda-\lambda_r \end{bmatrix} \rightarrow$$

The fundamental matrix $N = (I_2 - Q)^{-1} = \begin{bmatrix} 2\lambda & -2\lambda \\ -\lambda_r & \lambda+\lambda_r \end{bmatrix}^{-1} = \dfrac{1}{2\lambda^2}\begin{bmatrix} \lambda+\lambda_r & 2\lambda \\ \lambda_r & 2\lambda \end{bmatrix}$. Thus, given that the system starts in state 0, then the mean time to failure is given by $MTTF_0 =$

$\frac{(\lambda+\lambda_r)+2\lambda}{2\lambda^2} = \frac{3\lambda+\lambda_r}{2\lambda^2} = \frac{3\times0.015+0.5}{2(0.015)^2}$ = 1211.111hours, and $MTTF_1 = \frac{2\lambda+\lambda_r}{2\lambda^2}$ = 1177.77778 hours. The material on pages 101-104 of these notes may be used to compute the variance of TTF. It seems that for a 2-unit serial system we have to make both states (1, 2) absorbing so that the $MTTF_{Sys} = 1/(2\lambda) = 1/(0.03) = 33.3333$ hours.

Ternary Model For a Single Repairable Component

Consider a system that consists of a single component but the component can be in perfect (i.e., reliable) state 0, it can be in a derated (needing attention and or easier to repair) state 1, or in a completely failed state 2. As an example, a pump may deliver only a fraction of its rated output if one of its parts is not functioning perfectly. A multi-engine aircraft may experience problems with only one engine. The state TRD is given below. The state TRD Figure 10 shows that $(\lambda_0 + \lambda)\,\pi_0 = r_1\,\pi_1 + r\,\pi_2$; $(\lambda_1 + r_1)\times\pi_1 = \lambda_0\pi_0 + r_2\,\pi_2$, and $\pi_0 + \pi_1 + \pi_2 = 1$. These lead to the following nonhomogeneous system of equations.

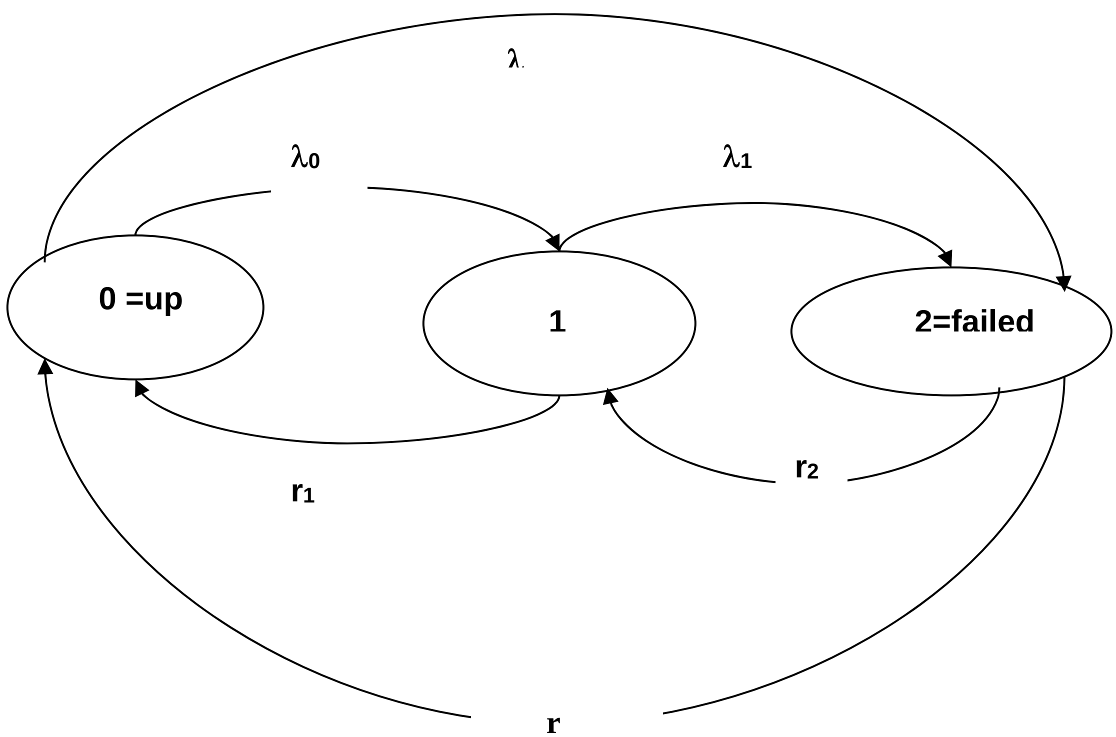


Figure 10. The TRD for a 3-state Repairable Component

$$\begin{cases}(\lambda_0+\lambda)\pi_0 \quad -r_1\pi_1 - r\,\pi_2 = 0\\ -\lambda_0\pi_0 + (r_1+\lambda_1)\pi_1 - r_2\,\pi_2 = 0\\ \pi_0 \quad + \pi_1 \quad + \pi_2 = 1\end{cases}$$ $\rightarrow$ **The determinants of this system are given by**

$$|A| = \begin{vmatrix}(\lambda_0+\lambda) & -r_1 & -r\\ -\lambda_0 & (r_1+\lambda_1) & -r_2\\ 1 & 1 & 1\end{vmatrix} = \lambda(\lambda_1 + r_1 + r_2) + \lambda_0(\lambda_1 + r_2 + r) + r(\lambda_1 + r_1) + r_1r_2,\ |A_0| =$$

$$= \begin{vmatrix}0 & -r_1 & -r\\ 0 & (r_1+\lambda_1) & -r_2\\ 1 & 1 & 1\end{vmatrix} = r(r_1+\lambda_1) + r_1r_2, \text{ and } |A_1| = \begin{vmatrix}(\lambda_0+\lambda) & 0 & -r\\ -\lambda_0 & 0 & -r_2\\ 1 & 1 & 1\end{vmatrix} = (\lambda_0+\lambda)r_2 + r\,\lambda_0.$$

Thus, the system mean availability in the long-term is given by $A_I = \pi_0 + \pi_1 = \dfrac{|A_0|+|A_1|}{|A|}$. For example, if $\lambda_0 = 0.01$ per hour, $\lambda_1 = 0.02$, $\lambda = 0.005$, $r_1 = 0.55$, $r_2 = 0.45$, and $r = 0.35$ per hour, then A $= \dfrac{|A_0|+|A_1|}{|A|} = \dfrac{0.4470+0.01025}{0.46030} = 0.9933738866$, and the outage rate U = 0.0066261134.

Again to compute the MTTF given that we are either in state 0 or 1, we have to assume that state 2 is absorbing and hence our one-step (= Δt = 1 unit) transition pr matrix is given by

$$P = \begin{array}{c} \\ 2\\ 0\\ 1\end{array}\begin{array}{ccc} 2 & 0 & 1\\ \end{array}\begin{bmatrix}1 & 0 & 0\\ \lambda & 1-\lambda-\lambda_0 & \lambda_0\\ \lambda_1 & r_1 & 1-\lambda_1-r_1\end{bmatrix} \rightarrow Q = \begin{bmatrix}1-\lambda-\lambda_0 & \lambda_0\\ r_1 & 1-\lambda_0-r_1\end{bmatrix}$$ $\rightarrow$ The fundamental matrix

$$N = (I_2 - Q)^{-1} = \begin{bmatrix}\lambda+\lambda_0 & -\lambda_0\\ -r_1 & \lambda_1+r_1\end{bmatrix}^{-1} = \frac{1}{\lambda_1\lambda_0 + \lambda\lambda_1 + \lambda r_1}\begin{bmatrix}\lambda_1+r_1 & \lambda_0\\ r_1 & \lambda+\lambda_0\end{bmatrix}.$$ Thus,

given that the system starts in state 0, then the mean time to failure is given by $MTTF_0$ =

$\dfrac{(\lambda_1 + r_1) + \lambda_0}{\lambda_1\lambda_0 + \lambda\lambda_1 + \lambda r_1}$. For example, if λ_0 = 0.01 per hour, λ_1 = 0.02, λ = 0.005, r_1 = 0.55 per hour, then

$\text{MTTF}_0 = \dfrac{(\lambda_1 + r_1) + \lambda_0}{\lambda_1\lambda_0 + \lambda\lambda_1 + \lambda r_1} = \dfrac{\mathbf{0.03 + 0.55}}{\mathbf{0.0002 + 0.0001 + 0.00275}}$ = 190.16394, and similarly, $\text{MTTF}_1 =$

$\dfrac{\lambda + \lambda_0 + r_1}{\lambda_1\lambda_0 + \lambda\lambda_1 + \lambda r_1} = \dfrac{\mathbf{0.015 + 0.55}}{\mathbf{0.0002 + 0.0001 + 0.002750}}$ = 185.24590 hours. The material on pages 109-110 of these notes may be used to compute the variance of TTF. Something is amiss in the above MTT failures because it seems that maintenance has not much added to the length of MTTF due to the fact that the MTTF W/O maintenance is roughly $1/\lambda$ = 200 > 190.16394. I am not certain where my error in the above Markov analysis is, but a glance at the TRD verifies the fact that my Markov analysis ignores the repair rates r_2 and r. In order to incorporate these two maintenance rates in computing the MTTF, it seems that there is no option but to create a 3rd auxiliary state 3 wherein the unit is no longer repairable and has to be replaced with a brand new unit. Thus state 3 is now in fact absorbing because the transitions $3 \rightarrow i$ (i = 0, 1, or 2) are no longer accessible. Let us assume that the transition rate from state 2 to 3 is given by, say λ_2 = 0.025, and hence

$$\mathbf{P} = \begin{array}{c} \\ 3 \\ 0 \\ 1 \\ 2 \end{array}\begin{array}{c} \begin{array}{cccc} 3 & 0 & 1 & 2 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-\lambda-\lambda_0 & \lambda_0 & \lambda \\ 0 & r_1 & 1-\lambda_1-r_1 & \lambda_1 \\ \lambda_2 & r & r_2 & 1-\lambda_2-r-r_2 \end{bmatrix} \end{array} \rightarrow \mathbf{Q} = \begin{bmatrix} 1-\lambda-\lambda_0 & \lambda_0 & \lambda \\ r_1 & 1-\lambda_1-r_1 & \lambda_1 \\ r & r_2 & 1-\lambda_2-r_2-r \end{bmatrix} \rightarrow$$

The fundamental matrix for our specific example with λ_0 = 0.01 per hour, λ_1 = 0.02, λ = 0.005, r_1 = 0.55, r_2 = 0.45, r = 0.35 and λ_2 = 0.025 per hour becomes $N = (I_3 - Q)^{-1}$ =

$$\begin{bmatrix} 0.9850 & 0.0100 & 0.0050 \\ 0.5500 & 0.4300 & 0.0200 \\ 0.3500 & 0.4500 & 0.1750 \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} 6049.18033 & 137.70492 & 40 \\ 6042.62295 & 139.34426 & 40 \\ 5862.29508 & 134.42623 & 40 \end{bmatrix}$$ $\rightarrow \text{MTTF}_0 = \mathbf{6049.18033 + 137.70492 + 40}$ = 6226.885245 hours, and

$\text{MTTF}_1 = \mathbf{6042.62295 + 139.34426 + 40}$ = 6221.967213 hours. These MTTF's make more sense than the previous ones as the component is repairable, and thus its MTTF that includes repair should exceed that of the case W/O restoration.

Series-Parallel System with Restoration

Consider a series-parallel system with failure rate λ_a, λ_b, and λ_c, where units A & B are in parallel redundancy in series with unit C, and from a design standpoint we assume that $\lambda_c < \lambda_a$ and λ_b. We first apply the Markov-Model approach to obtain the reliability function at time t assuming that all 3 components are irreparable (or non-restorable). To this end, we define 3 state spaces: State 1 = all 3 units are actively reliable, State 2 = only B & C are up and A is down (or failed), State 3 = only A & C are up and B is down, and State 4 = System has failed, which is an absorbing state. The TRD in Figure 11 shows that

$$P_1(t+\Delta t) = P_1(t)(1 - \lambda_c\Delta t)(1 - \lambda_b\Delta t)(1 - \lambda_a\Delta t)$$

$$P_2(t+\Delta t) = P_1(t)\,(\lambda_a\Delta t) + P_2(t)\,(1 - \lambda_b\Delta t)\,(1 - \lambda_c\Delta t) \qquad (88)$$

$$P_3(t+\Delta t) = P_1(t)\,(\lambda_b\Delta t) + P_3(t)\,(1 - \lambda_a\Delta t)\,(1 - \lambda_c\Delta t)$$

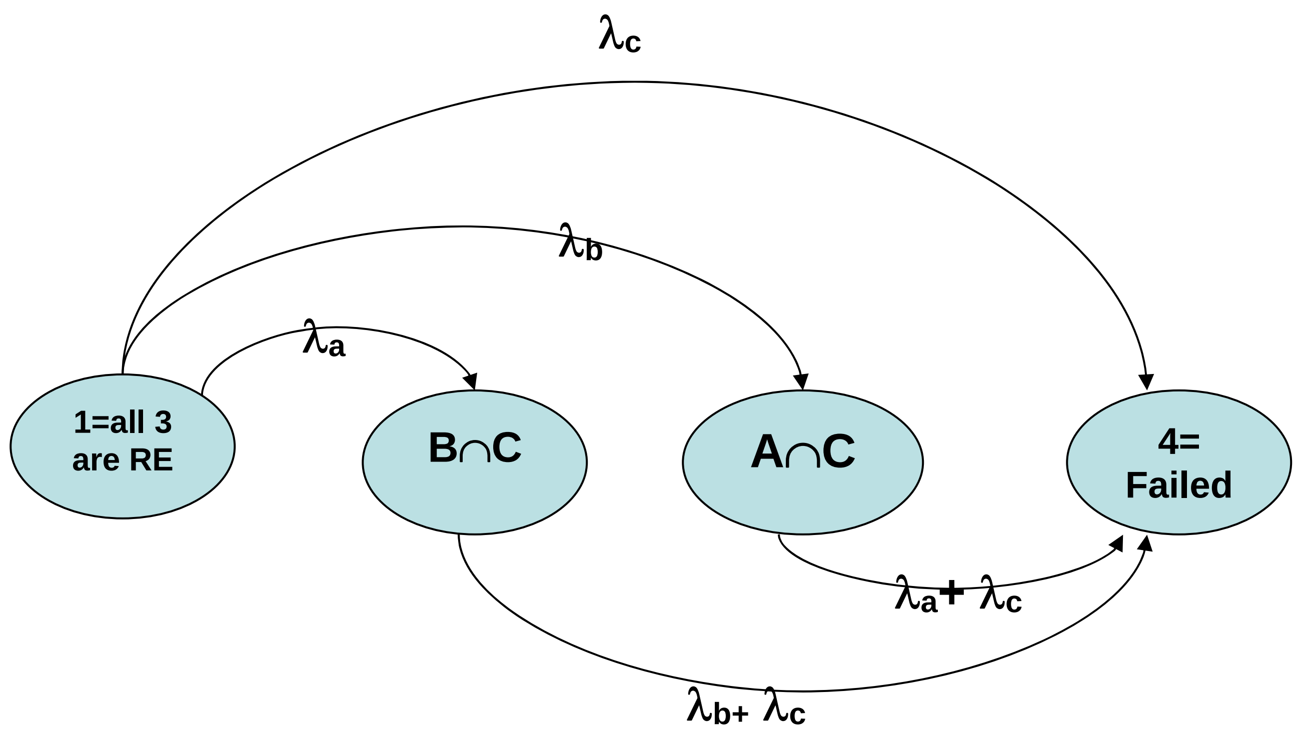


Figure 11. The TRD for a 3-unit S-P non-restorable System

Note that the Pr equation for finding the system in the state 4 = "Failed" is duplicative to the above 3 equations because $R(t) = P_1(t) + P_2(t) + P_3(t) = 1 - F(t)$. Transposing $P_i(t)$, i = 1, 2, 3 to the LHS of equation (95), dividing by Δt, and taking the limit as $\Delta t \rightarrow 0$ results in the following system of ddes.

$$\begin{cases} \mathbf{P_1'(t) = -(\lambda_c + \lambda_b + \lambda_a)P_1(t)} \\ \mathbf{P_2'(t) = \lambda_a P_1(t) - (\lambda_b + \lambda_c)P_2(t)} \\ \mathbf{P_3'(t) = \lambda_b P_1(t) - (\lambda_c + \lambda_a)P_3(t)} \end{cases}$$

The above system can be easily solved for any failure rates λ, but to facilitate understanding of the procedure, we assume that $\lambda_a = \lambda_b = 0.0001$ and $\lambda_c = 0.00005$. Inserting these into the above system of ddes yields

$$\begin{cases} \frac{dP_1(t)}{dt} = -0.00025\,P_1(t) \\ \frac{dP_2(t)}{dt} = 0.0001\,P_1(t) - 0.00015\,P_2(t) \\ \frac{dP_3(t)}{dt} = 0.0001\,P_1(t) - 0.00015\,P_3(t) \end{cases} \rightarrow \quad \frac{dP_1(t)}{P_1(t)} = -0.00025\,dt \rightarrow$$

$\ln[P_1(t)] = -0.00025\,t + \ln C \rightarrow P_1(t) = \mathbf{e^{-0.00025t + \ln C}} \rightarrow P_1(t) = \mathbf{C\,e^{-0.00025t}}$.

We now apply the boundary condition $P_1(t = 0) = 1$, $P_2(t = 0) = 0$, $P_3(t = 0) = 0$, and

$P_F(t = 0) = P_4(t = 0) = 0$. This gives C = 1 so that $P_1(t) = \mathbf{e^{-0.00025t}}$. Inserting this function into the 2nd dde yields $\frac{dP_2(t)}{dt} = 0.0001\,e^{-0.00025\,t} - 0.00015\,P_2(t) \rightarrow \frac{dP_2(t)}{dt} + 0.00015\,P_2(t) = 0.0001\,e^{-0.00025\,t}$. This last is a first-order differential equation with an integrating factor (IF) equal to $e^{0.00015\,t}$. Multiplying both sides by this IF yields $\frac{dP_2(t)}{dt}e^{0.00015t} + 0.00015\,e^{0.00015t}P_2(t) = 0.0001\,e^{-0.0001t} \rightarrow$

$\frac{d}{dt}[P_2(t)e^{0.00015t}] = 0.0001\,e^{-0.0001t} \rightarrow P_2(t)e^{0.00015t} = -e^{-0.0001t} + C \rightarrow P_2(t) = -e^{-0.00025\,t} + C\,e^{-0.00015\,t}$. Applying the boundary condition $P_2(t = 0) = 0$ yields C = 1. Hence, $P_2(t) = e^{-0.00015t}$

$-\mathbf{e}^{-0.00025t}$. Similar procedure yields $P_3(t) = e^{-0.00015t} - \mathbf{e}^{-0.00025t}$. Hence, $R(t) = P_1(t) + P_2(t) + P_3(t) = 2e^{-0.00015t} - \mathbf{e}^{-0.00025t}$. For example, the system reliability for a 2000-hr mission is equal to $R_{Sys}(2000) = 2e^{-0.30} - \mathbf{e}^{-0.5} = 0.875105782$. To verify that the above Markov-Model procedure to system reliability has given the correct result for this non-restorable system, we clearly observe that the system is reliable iff both subsystems 1 (consisting of the pure-parallel units A and B) & component C are reliable. The reliability of the pure-parallel subsystem 1 is given by $R_1(t) = 1 - (1 - \mathbf{e}^{-0.0001t})^2 = 2\mathbf{e}^{-0.0001t} - \mathbf{e}^{-0.0002t}$. Hence, $R_{Sys}(t) = (2\mathbf{e}^{-0.0001t} - \mathbf{e}^{-0.0002t})\,\mathbf{e}^{-\lambda_c t} = (2\mathbf{e}^{-0.0001t} - \mathbf{e}^{-0.0002t})\,\mathbf{e}^{-0.00005t} = 2e^{-0.00015t} - \mathbf{e}^{-0.00025t}$ as before! Applying the Markov procedure to an irreparable system, as you can see, is totally unnecessary and was done here only for illustrative purposes. Because, obtaining the $R_{Sys}(t)$ for a repairable system is far more complicated and requires the use of Markovian procedure, which I will do after finishing the analysis of our non-repairable system. From the reliability function the MTTF for the non-restorable system is given by $MTTF_{Sys} = 2/0.00015 - 1/0.00025 = 9333.333333$ hours. Since the this system is irreparable, then its availability at time t is given by A(t) = R(t) (it may be even more appropriate to argue that the definition of availability is not applicable to non-restorable systems), and its dependability is given by D(t) = R(t); thus D(2000 hours) = 87.5105782%.

We now add on-line restoration to the above system by assuming that when unit A fails it can be restored at the rate of $r_a = 0.40$/hr and similarly when B fails it can be restored at the rate $r_b = 0.40$ per hour. Note that when unit C fails the entire system fails, and thus we have to assume that when the system is in the failed state (which can also happen by combinations AC down, BC down, and all 3 down), two units can be repaired on-line at the same time by 2 different servers at either the rate $r_a + r_c$, or $r_b + r_c$. I must alert you to the fact that this assumption violates that of a B&D process because two transitions (i.e., two repairs) during Δt is disallowed. We are making this assumption for the sake of simplicity and have to further assume that when only unit C is down, then because there are two servers C will get repaired at the rate of 2r, where $r_a = r_c = r_b = r$, and $\lambda_a = \lambda_b = \lambda$. Therefore, the following analysis will lead to only a rough approximation to the exact solution having 8 different states. Then, the TRD 11 modifies to the TRD given in Figure 12. From Figure 12 we obtain

$$\begin{cases} P_1(t+\Delta t) = P_1(t)(1-\lambda_c\Delta t - 2\lambda\Delta t) + P_2(t)r\Delta t + P_3(t)r\Delta t + P_4(t)\,2r\Delta t \\ P_2(t+\Delta t) = P_1(t)(\lambda\Delta t) + P_2(t)(1-r\Delta t)(1-\lambda\Delta t-\lambda_c\Delta t) + P_4(t)(2r)\Delta t \\ P_3(t+\Delta t) = P_1(t)(\lambda\Delta t) + P_3(t)(1-r\Delta t)(1-\lambda_c\Delta t-\lambda\Delta t) + P_4(t)(2r)\Delta t \\ P_4(t+\Delta t) = P_1(t)(\lambda_c\Delta t) + P_2(t)(\lambda_c\Delta t+\lambda\Delta t) + P_3(t)(\lambda_c\Delta t+\lambda\Delta t)P_4(t)(1-4r\Delta t) \end{cases}$$

The above system leads to the following system of ddes which is given below Figure 12, where for simplicity we are assuming $r = r_a = r_c = r_b$, and $\lambda = \lambda_a = \lambda_b$.

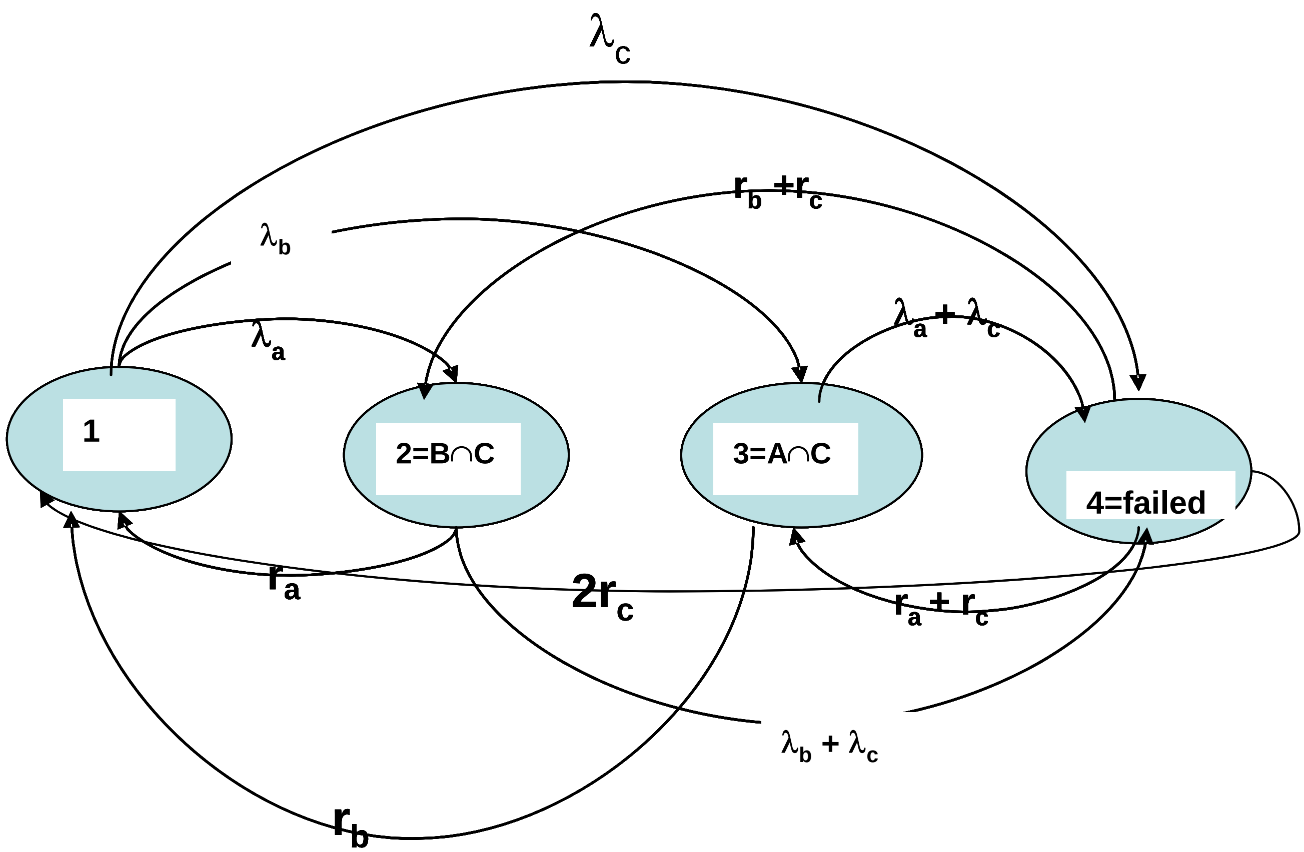


Figure 12. The TRD for a 3-unit S-P On-line restorable System

$$\begin{cases} \mathbf{P_1'(t) = P_1(t)(-\lambda_c - 2\lambda) + P_2(t)r + P_3(t)r + 2rP_4(t)} \\ \mathbf{P_2'(t) = P_1(t)\lambda + P_2(t)(-r - \lambda - \lambda_c) + 2rP_4(t)} \\ \mathbf{P_3'(t) = P_1(t)\lambda + P_3(t)(-r - \lambda_c - \lambda) + 2rP_4(t)} \\ \mathbf{P_4'(t) = P_1(t)\lambda_c + P_2(t)(\lambda_c + \lambda) + P_3(t)(\lambda + \lambda_c) - 6rP_4(t)} \end{cases} \rightarrow P'(t) = B\,P(t), \text{ where the}$$

Note that I have done a complete Markov analysis of the renewal system under S-P On-Line System which is listed on my website.

TRM (transition rate matrix) is given by $B = \begin{bmatrix} -(2\lambda+\lambda_c) & r & r & 2r \\ \lambda & -(r+\lambda+\lambda_c) & 0 & 2r \\ \lambda & 0 & -(r+\lambda+\lambda_c) & 2r \\ \lambda_c & \lambda+\lambda_c & \lambda+\lambda_c & -6r \end{bmatrix}$.

Obtaining the most general transient (i.e., time-dependent) solution to this last system of ddes is not an easy task (I am not sure that is worth the time and effort required). However, before obtaining the steady-state solution that will yield the long-term availability $A = \pi_1 + \pi_2 + \pi_3$, we may obtain the MTBF by obtaining the fundamental matrix $N = (I_3 - Q)^{-1}$. In order to obtain N, we must obtain the transition pr matrix, which is given below for the case $r_a = r_b = r_c = r$.

$$\mathbf{P} = \begin{array}{c} \\ 4 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 4 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1-6r & 2r & 2r & 2r \\ \lambda_c & 1-(2\lambda+\lambda_c) & \lambda & \lambda \\ \lambda+\lambda_c & r & 1-(r+\lambda+\lambda_c) & 0 \\ \lambda+\lambda_c & r & 0 & 1-(r+\lambda+\lambda_c) \end{bmatrix} \end{array}$$

Since $dR_{Sys}(t)/dt = \frac{d}{dt}[P_1(t) + P_2(t) + P_3(t)] = P_1'(t) + P_2'(t) + P_3'(t) = -P_4'(t) = -\lambda_c P_1(t) - (\lambda + \lambda_c)P_2(t) - (\lambda + \lambda_c)P_3(t) + 6rP_4(t)$, then making the location 4 = "Failed" = { $\bar{A}\bar{B}$, $\bar{C}$, $\bar{A}\bar{C}$, $\bar{B}\bar{C}$, $\bar{A}\bar{B}\bar{C}$ }, an absorbing state, we obtain the one-step 4×4 canonical transition Pr matrix as shown below.

$$\tilde{\mathbf{P}} = \begin{array}{c} \\ 4 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 4 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_c & 1-(2\lambda+\lambda_c) & \lambda & \lambda \\ \lambda+\lambda_c & r & 1-(r+\lambda+\lambda_c) & 0 \\ \lambda+\lambda_c & r & 0 & 1-(r+\lambda+\lambda_c) \end{bmatrix} \end{array}$$

Note that $\frac{d}{dt}[P_1(t)+P_2(t)+P_3(t)+P_4(t)] = \frac{d}{dt}[1] = 0$ shows that once we have information about states 1, 2, and 3, then we also have complete information on state 4. The truncated matrix Q corresponding to the above stochastic matrix is given by

$$\mathbf{Q} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1-(2\lambda+\lambda_c) & \lambda & \lambda \\ r & 1-(r+\lambda+\lambda_c) & 0 \\ r & 0 & 1-(r+\lambda+\lambda_c) \end{bmatrix} \end{array} \quad \rightarrow$$

$$\mathbf{I_3 - Q} = \begin{bmatrix} (\lambda_a+\lambda_b+\lambda_c) & -\lambda_a & -\lambda_b \\ -r_a & (r_a+\lambda_b+\lambda_c) & 0 \\ -r_b & 0 & (r_b+\lambda_a+\lambda_c) \end{bmatrix}.$$

Obtaining the inverse of $(I_3 - Q)$ for the general case is at best very messy, although it can be done using determinants and adjoint matrices, which is the transpose of matrix C consisting of cofactors. However, we limit discussion to the case $\lambda_a = \lambda_b = 0.0001$, $\lambda_c = 0.00005$ and $r = r_a = r_b = 0.40$ per hour. Then, the fundamental matrix is given by

$$\mathbf{N} = (\mathbf{I_3 - Q})^{-1} = \begin{bmatrix} 0.00025 & -0.0001 & -0.0001 \\ -0.40 & 0.40115 & 0 \\ -0.40 & 0 & 0.40015 \end{bmatrix}^{-1} = \begin{bmatrix} 19970.05615 & 0.00050 & 0.00050 \\ 19962.57018 & 0.00075 & 0.00050 \\ 19962.57018 & 0.00050 & 0.00075 \end{bmatrix}$$

The above matrix N shows that given we are in state 1, then the $MTBF_{Sys}$ is equal to $MTBF_{Sys1}$ = 19970.05615 + 0.00050 + 0.00050 = 19970.05715 hours, but when we are in state 2 or 3, then the $MTBF_{Sys2(3)}$ = 19962.57018 + 0.00125 = 19962.57143 hours. Thus, restoration has more than doubled our $MTBF_{Sys}$ from 9333.33333 hours to roughly 19970 hours. To obtain the steady-state solutions, the TRD 12 shows that $(2\lambda + \lambda_c)\pi_1 = r\,\pi_2 + r\,\pi_3$, $\lambda\,\pi_1 + 2r\pi_4 = (r + \lambda + \lambda_c)\pi_2$, $\lambda\,\pi_1 + 2r\pi_4 = (r+\lambda + \lambda_c)\pi_3$, and $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$. Solving this system of 4 equations with 4 unknowns simultaneously for the specific values of $\lambda = 0.0001$, $\lambda_c = 0.00005$ and $\mu = 0.40$ yields $\pi_1 = 0.99934412186471$, $\pi_2 = \pi_3 = 0.00031229503808$, and $\pi_4 = 0.00003128805913$. Thus the value of long-term system availability is given by $A_I = \pi_1 + \pi_2 + \pi_3 = 0.99996871194087$, and its long-term average $MTBF_{Sys} = 19970.05715 \times 0.99934412186471 + 19962.57143 \times 2 \times 0.00031229503808 + 0 \times 0.00003128805913 = 19969.4276502$ hours, using the fact that $MTTF_{Sys(4)} = 0$.

The 2-Unit Pure Parallel System with no On-Line Restoration but with System Restoration Rate r (or λ_r)

Since there is no on-line restoration, then the entire system gets repaired after both units have failed, and as a result μ represents the rate at which the system gets restored. We can define the sates as either X(t) = number of operational units, or number of units that are down (it makes no difference as long as one is consistent with their definition). To be consistent, let X(t) = number of units that **are** down or in the failed state. Then, $R_x = \{0, 1, 2\}$. The TRD for the system is given in figure 13.

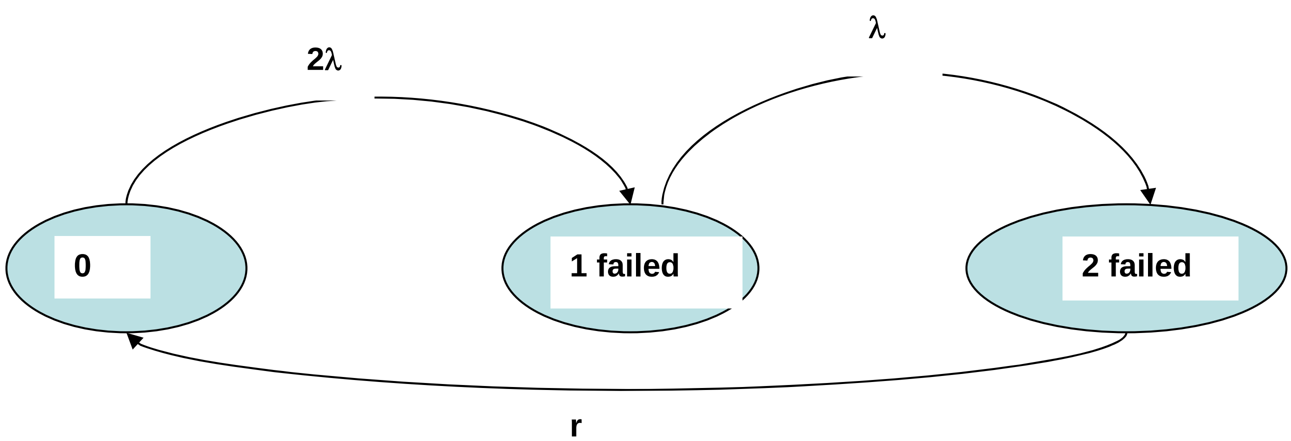


Figure 13. 2-uint Pure Parallel with Off-line Restoration

Figure 13 clearly shows that $2\lambda\pi_0 = r\,\pi_2$ and $2\lambda\,\pi_0 = \lambda\,\pi_1$, and using the fact that $\pi_0 + \pi_1 + \pi_2 \equiv 1$, we obtain $\pi_0 = \dfrac{r}{3r + 2\lambda}$, $\pi_1 = \dfrac{2r}{3r + 2\lambda}$ and hence system availability is given by $A = \pi_0 + \pi_1 =$

$\frac{\mathbf{3r}}{\mathbf{3r+2\lambda}}$. It can be easily shown that $MTBF_0 = 3/(2\lambda)$ and $MTBF_1 = 1/\lambda$.

The 2-out-of-3 System with Off-Line Restoration Rate λ_r

Again we define 0 as no failures, state 1 as exactly one unit failed, and state D = 2 or more units failed leading to system being down. The TRD is given in figure 14. As in the case of the 2-unit Parallel system it can easily be shown from Figure 14 that the system availability is given by $A_I = \pi_0 + \pi_1 = \frac{2\lambda_r}{5\lambda_r + 6\lambda} + \frac{3\lambda_r}{5\lambda_r + 6\lambda} = \frac{5\lambda_r}{5\lambda_r + 6\lambda}$.

<u>Exercise24.</u> Specifically consider a 2-out-of-3 redundant system with on-line restoration of r per time unit and only one restoration at a time, i.e., only one server or repair facility. By on-line restoration we mean that system operations will not be totally interrupted while repair is ongoing. The TRD diagram is given below. Let X_t represent the number of failed units at time t, where one step is equal to Δt.

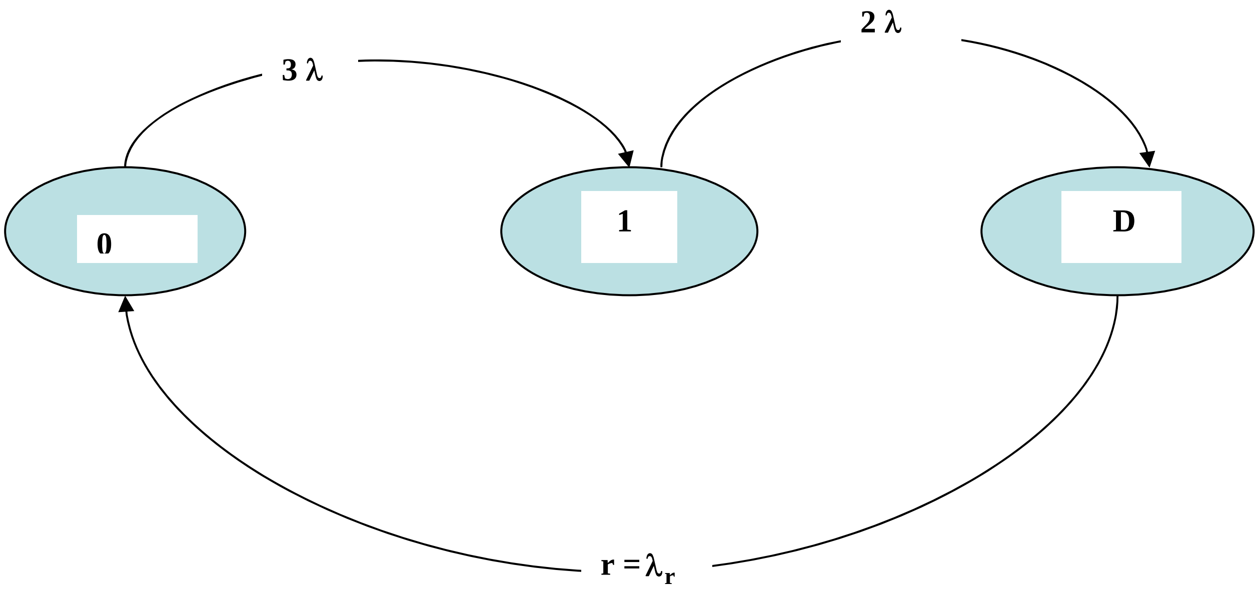


Figure 14. The TRD for the 2-out-of-3 System with off-line Restoration

From the TRD 15 we can easily write the following system of ddes

$$\begin{cases} P_0'(t) = P_1(t)(-3\lambda) + P_1(t)\lambda_r \\ P_1'(t) = P_0(t)(3\lambda) + \lambda_r P_2(t) - (2\lambda + \lambda_r)P_1(t) \\ P_2'(t) = 2\lambda P_1(t) + \lambda_r P_3(t) - (\lambda + \lambda_r)P_2(t) \\ P_3'(t) = \lambda P_2(t) - \lambda_r P_3(t) \end{cases}$$

$\rightarrow$ $P'(t) = B\,P(t)$, where the TRM is given by $B =$

$$\begin{bmatrix} -3\lambda & \lambda_r & 0 & 0 \\ 3\lambda & -(\lambda_r + 2\lambda) & \lambda_r & 0 \\ 0 & 2\lambda & -(\lambda + \lambda_r) & \lambda_r \\ 0 & 0 & \lambda & -\lambda_r \end{bmatrix}$$

, and hence the time-dependent solution is given by $P(t) = M\, e^{\Lambda t} M^{-1}$, where

$$\Lambda = \begin{bmatrix} \xi_1 & 0 & 0 & 0 \\ 0 & \xi_2 & 0 & 0 \\ 0 & 0 & \xi_3 & 0 \\ 0 & 0 & 0 & \xi_4 \end{bmatrix}.$$

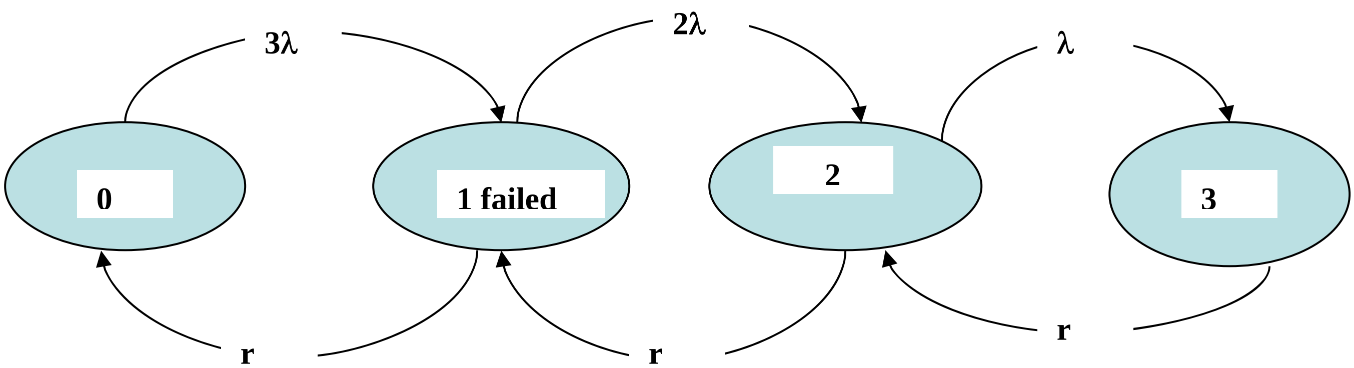


Figure 15. The TRD for a 2-out-of-3 System with on-line restoration

Use the TRD 15 to show that $\pi_0 = \dfrac{1}{1+(3\lambda/\lambda_r)+6(\lambda/\lambda_r)^2+6(\lambda/\lambda_r)^3} = \dfrac{\lambda_r^3}{\text{Den}}$, where Den $=\lambda_r^3$ $+ 3\lambda\lambda_r^2 + 6\lambda^2\lambda_r + 6\lambda^3$. Further, $\pi_1 = (3\lambda/\lambda_r)\pi_0 = \dfrac{3\lambda\lambda_r^2}{\text{Den}}$. Thus, the average availability in the long-run is given by $A_I = \pi_0 + \pi_1 = \dfrac{\lambda_r^3+3\lambda\lambda_r^2}{\lambda_r^3+3\lambda\lambda_r^2+6\lambda^2\lambda_r+6\lambda^3}$. Further, obtain the $\text{MTBF}_0 = \dfrac{\lambda_r+5\lambda}{6\lambda^2}$ and $\text{MTBF}_1 = \dfrac{\lambda_r+3\lambda}{6\lambda^2}$.

Table C.16 on Page 384 of Paul Kales (*Reliability For Technology, Engineering, and Management*, Prentice Hall, ISBN:0-13-485822-0)

Paul Kales (1998) tabulates the results of k-out-n standby redundant systems with on-line restoration and exactly one repair at a time. I will specifically work out the analysis for a 3-unit standby system where one unit is needed for operational success at any time t. In standby systems it is generally best to define X_t to represent the number of failed units at time t (instead of number of operating units at time t). The TRD is given in figure 16.

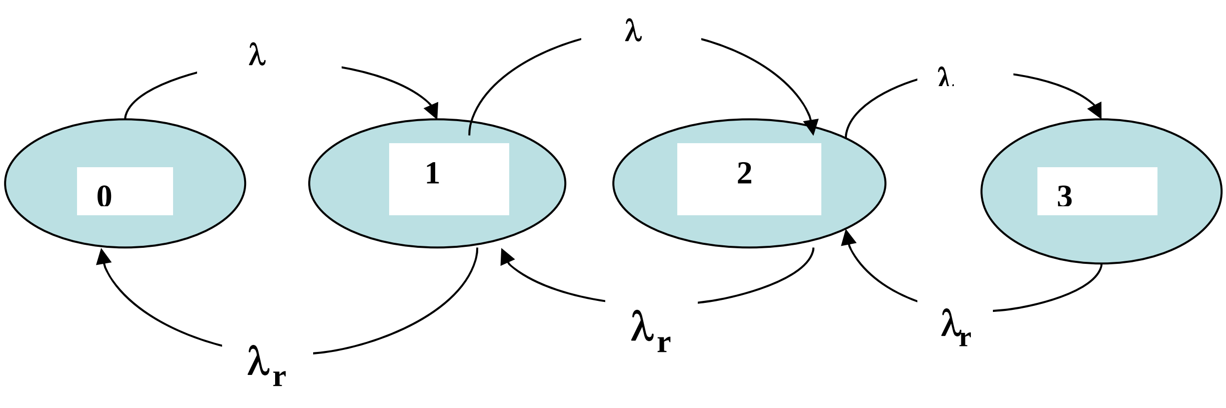


Figure 16. The TRD for a 3-unit repairable Standby System

The TRD 16 shows that $\lambda\,\pi_0 = \lambda_r\,\pi_1 \rightarrow \pi_1 = (\lambda/\lambda_r)\pi_0$; $(\lambda + \lambda_r)\pi_1 = \lambda\,\pi_0 + \lambda_r\pi_2 \rightarrow$

$(\lambda + \lambda_r)\,\dfrac{\lambda}{\lambda_r}\pi_0 = \lambda\,\pi_0 + \lambda_r\pi_2 \rightarrow \dfrac{\lambda^2}{\lambda_r}\pi_0 = \lambda_r\pi_2 \rightarrow \pi_2 = (\lambda^2/\lambda_r^2)\pi_0$. In a similar fashion $\pi_3 =$

$(\lambda^3/\lambda_r^3)\pi_0 \rightarrow \pi_0 + (\lambda/\lambda_r)\pi_0 + (\lambda/\lambda_r)^2\pi_0 + (\lambda/\lambda_r)^3\pi_0 = 1 \rightarrow$

$\pi_0 = \dfrac{1}{1+(\lambda/\lambda_r)+(\lambda/\lambda_r)^2+(\lambda/\lambda_r)^3} = \dfrac{\lambda_r^3}{\mathbf{Den}}$, where Den = $\lambda_r^3 + \lambda\lambda_r^2 + \lambda^2\lambda_r + \lambda^3$.

Further, $\pi_1 = (\lambda/\lambda_r)\pi_0 = \dfrac{\lambda\lambda_r^2}{\mathbf{Den}}$. Thus, the average availability in the long-run is given by

$A_I = \pi_0 + \pi_1 + \pi_2 = \dfrac{\lambda_r^3+\lambda\lambda_r^2+\lambda^2\lambda_r}{\lambda_r^3+\lambda\lambda_r^2+\lambda^2\lambda_r+\lambda^3}$, which is identical to row 4 of Table C.16 on page 384 of Paul Kales. Therefore, the system outage rate is given by $U = 1 - A_I =$

$\dfrac{\lambda^3}{\lambda_r^3+\lambda\lambda_r^2+\lambda^2\lambda_r+\lambda^3}$. Although, Kales does not provide the $MTBF_{Sys}$, I will work out the details below.

To obtain the $MTBF_{Sys}$ of the above 3-unit standby system, we must first write the one-step (= Δt) transitional pr matrix, P.

$$\mathbf{P} = \begin{bmatrix} 1-\lambda & \lambda & 0 & 0 \\ \lambda_r & 1-(\lambda_r+\lambda) & \lambda & 0 \\ 0 & \lambda_r & 1-(\lambda+\lambda_r) & \lambda \\ 0 & 0 & \lambda_r & 1-\lambda_r \end{bmatrix}.$$

Since state {3} is the only failed (or down = {D}) state, then we will make it absorbing. This leads to the following matrix in canonical form.

$$\tilde{\mathbf{P}} = \begin{matrix} \\ D \\ 0 \\ 1 \\ 2 \end{matrix} \begin{matrix} \begin{matrix} D & 0 & 1 & 2 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-\lambda & \lambda & 0 \\ 0 & \lambda_r & 1-\lambda_r-\lambda & \lambda \\ \lambda & 0 & \lambda_r & 1-\lambda_r-\lambda \end{bmatrix} \end{matrix} \rightarrow \mathbf{Q} = \begin{bmatrix} 1-\lambda & \lambda & 0 \\ \lambda_r & 1-\lambda_r-\lambda & \lambda \\ 0 & \lambda_r & 1-\lambda_r-\lambda \end{bmatrix}$$

$\rightarrow$

$$(\mathbf{I_3} - \mathbf{Q}) = \begin{bmatrix} \lambda & -\lambda & 0 \\ -\lambda_r & \lambda_r + \lambda & -\lambda \\ 0 & -\lambda_r & \lambda_r + \lambda \end{bmatrix}.$$ Thus the fundamental matrix is given by

$$\mathbf{N} = (\mathbf{I_3} - \mathbf{Q})^{-1} = \frac{\mathrm{Adj}(I_3 - Q)}{\det(I_3 - Q)} = \frac{\begin{bmatrix} \lambda^2 + r^2 + \lambda r & \lambda(\lambda + \lambda_r) & \lambda^2 \\ \lambda_r(\lambda + \lambda_r) & \lambda(\lambda + \lambda_r) & \lambda^2 \\ r^2 & \lambda\lambda_r & \lambda^2 \end{bmatrix}}{\lambda^3}.$$

This last fundamental matrix shows that $MTBF_0 = \dfrac{r^2 + 3\lambda^2 + 2\lambda\lambda_r}{\lambda^3}$, $MTBF_1 = \dfrac{r^2 + 2\lambda^2 + 2\lambda\lambda_r}{\lambda^3}$, and $MTBF_2 = \dfrac{r^2 + \lambda^2 + \lambda\lambda_r}{\lambda^3}$, where $r^2 = \lambda_r^2$.

Examples 10.15 Through 10.18 on Pages 223-226 of Paul Kales (1998)

From the Figure 17 (the state-TRD) on my website, the equilibrium equation from node 1 is obtained by equating the flow rate into node 1 to the flow rate out node 1.

$r_a\,\pi_2 + r_b\,\pi_3 + r_c\,\pi_4 = (\lambda_a + \lambda_b + \lambda_c)\,\pi_1 \rightarrow\ 0.000032\pi_1 - \pi_2 - 2\pi_3 - \pi_4 = 0.$

Node 2: $\lambda_a\,\pi_1 + r_c\,\pi_6 = (r_a + \lambda_b + \lambda_c)\,\pi_2 \rightarrow\ 0.000005\pi_1 - 1.000027\,\pi_2 + \pi_6 = 0.$

Node 3: $\lambda_b\,\pi_1 + r_a\,\pi_5 + r_c\,\pi_7 = (\lambda_a + r_b + \lambda_c)\,\pi_3 \rightarrow$

$$0.000025\pi_1 - 2.000007\,\pi_3 + \pi_5 + \pi_7 = 0.$$

Node 4: $\lambda_c\,\pi_1 = (\lambda_a + r_c + \lambda_b)\,\pi_4 \rightarrow\ 0.000002\pi_1 - 1.000030\,\pi_4 = 0.$

Node 5: $\lambda_b\,\pi_2 + \lambda_a\,\pi_3 + r_c\,\pi_8 = (r_a + \lambda_c)\,\pi_5 \rightarrow$

$$0.000025\pi_2 + 0.000005\,\pi_3 - 1.000002\,\pi_5 + \pi_8 = 0.$$

Node 6: $\lambda_c\,\pi_2 + \lambda_a\,\pi_4 = (r_c + \lambda_b)\,\pi_6 \rightarrow$

$$0.000002\pi_2 + 0.000005\,\pi_4 - 1.000025\,\pi_6 = 0.$$

Node 7: $\lambda_c \pi_3 + \lambda_b \pi_4 = (r_c + \lambda_a) \pi_7 \quad \rightarrow$

$$0.000002\pi_3 + 0.000025\,\pi_4 - 1.000005\,\pi_7 = 0.$$

Constraint: $\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 + \pi_6 + \pi_7 + \pi_8 = 1.$

Note that if you write the equilibrium steady-state equation for node 8, it will not generate another independent equation and thus we could not obtain a unique solution unless we impose the above constraint. Solving the above system of 8 equations with 8 unknowns yields

$\pi' = [0.99998050018525 \quad 0.00000499978751 \quad 0.00001249984375 \quad 0.0000199990100$

$0.00000000018749 \quad 0.00000000002000 \quad 0.00000000007500 \quad 0.00000000000000].$

Thus, the long-term availability is given by $A_I = \pi_1 + \pi_2 + \pi_3 = 0.99999799981651$, which is consistent with Kales answer near the bottom of his page 226. The outage rate is given by

$U = 1 - 0.99999799981651 = 0.00000200018349494346 4.$

In order to obtain the $MTBF_{Sys}$, we let D = {4, 5, 6, 7, 8} and make this down state as absorbing resulting in

$$\tilde{\mathbf{P}} = \begin{matrix} & \mathbf{D} & \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{D} & 1 & 0 & 0 & 0 \\ \mathbf{1} & \lambda_c & 1-\lambda_a-\lambda_b-\lambda_c & \lambda_a & \lambda_b \\ \mathbf{2} & \lambda_b+\lambda_c & r_a & 1-r_a-\lambda_b-\lambda_c & 0 \\ \mathbf{3} & \lambda_a+\lambda_c & r_b & 0 & 1-r_b-\lambda_a-\lambda_c \end{matrix} \rightarrow$$

$$\tilde{\mathbf{P}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2\times10^{-6} & 0.999968 & 5\times10^{-6} & 25\times10^{-6} \\ 27\times10^{-6} & 1 & -27\times10^{-6} & 0 \\ 7\times10^{-6} & 2 & 0 & -1.000007 \end{bmatrix} \rightarrow$$

$$Q = \begin{bmatrix} \mathbf{0.999968} & \mathbf{5\times10^{-6}} & \mathbf{25\times10^{-6}} \\ \mathbf{1} & \mathbf{-27\times10^{-6}} & \mathbf{0} \\ \mathbf{2} & \mathbf{0} & \mathbf{-1.000007} \end{bmatrix} \rightarrow$$

$$(I_3 - Q) = \begin{bmatrix} 0.000032 & -5\times10^{-6} & -25\times10^{-6} \\ -1 & 1.000027 & 0 \\ -2 & 0 & 2.000007 \end{bmatrix} \rightarrow$$

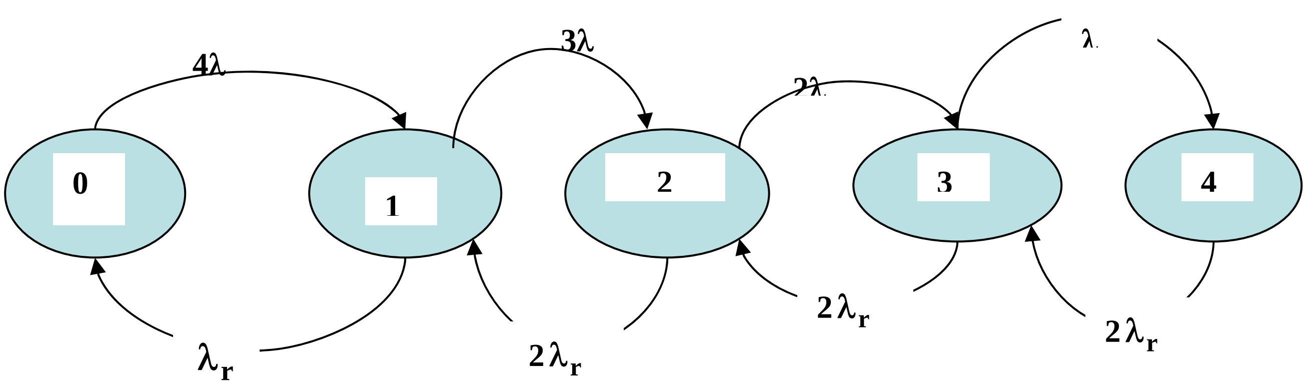
Thus, the fundamental matrix is given by

$$N = (I_3 - Q)^{-1} = \begin{bmatrix} \mathbf{4.99944382167162} & \mathbf{0.00002499654420} & \mathbf{0.00006249282905} \\ \mathbf{4.99930884033293} & \mathbf{0.00003499559932} & \mathbf{0.00006249114179} \\ \mathbf{4.99942632367949} & \mathbf{0.00002499645671} & \mathbf{0.00006749259282} \end{bmatrix} \times 10^5$$

$\rightarrow$

$MTBF_{Sys1}$ = 499953.131104487 hours, $MTBF_{Sys2}$ = 499940.632707404, and $MTBF_{Sys3}$ = 499951.881272903 hours.

Finally, the TRD for a 2-out-of-4 parallel system (i.e., all 4 units get energized at time zero) where two units can be repaired at a time is given by



Note that in the above diagram X_t = number of failed units at time t. Further, when we are in the state {2 failed}, the transition 2 → 0 during Δt has a Pr of $(2\lambda_r \Delta t)^2$ and thus in the limit as $\Delta t \rightarrow 0$ it drops out the C-K ddes.