Chapter 8
Markov Processes

Marvin Rausand
marvin.rausand@ntnu.no

RAMS Group
Department of Production and Quality Engineering
NTNU

(Version 0.1)

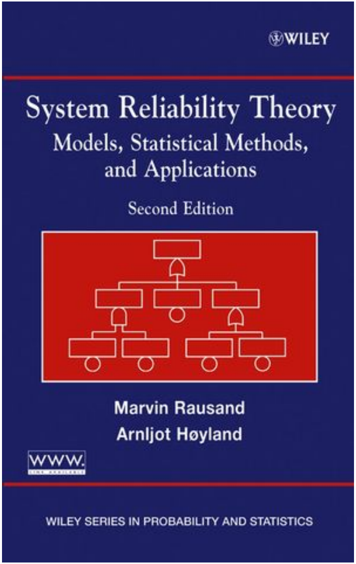
NTNU – Trondheim
Norwegian University of
Science and Technology

Slides related to the book

System Reliability Theory
Models, Statistical Methods, and Applications

Wiley, 2004

Homepage of the book:
http://www.ntnu.edu/ross/books/srt



Example 8.1

Consider a parallel structure of two components. Each component is assumed to have two states, a functioning state and a failed state. The structure has therefore $2^2 = 4$ possible states, and the state space is $\mathcal{X} = \{0, 1, 2, 3\}$

State	Component 1	Component 2
3	Functioning	Functioning
2	Functioning	Failed
1	Failed	Functioning
0	Failed	Failed

State space

Let $X(t)$ denote the state of the system at time t.

The state space is the set of all the possible system states. In this book we number the states by integers from 0 to r. The state space is therefore

$$\mathcal{X} = \{0, 1, 2, \ldots, r\}$$

Let $P_i(t) = \Pr(X(t) = i)$ be the probability that the system is in state i at time t.

The state probability distribution is denoted

$$\mathbf{P}(t) = (P_0(t), P_1(t), \ldots, P_r(t))$$

Markov property

Assume that the process is in state i at time s, that is, $X(s) = i$. The conditional probability that the process will be in state j at time $t + s$ is

$$\Pr(X(t + s) = j \mid X(s) = i, X(u) = x(u), 0 \leq u < s)$$

where $\{x(u), 0 \leq u < s\}$ denotes the "history" of the process up to, but not including, time s

The process is said to have the *Markov property* if

$$\begin{aligned}\Pr(X(t + s) = j \mid X(t) = i, X(u) = x(u), 0 \leq u < s)\\ = \Pr(X(t + s) = j \mid X(s) = i)\\ \text{for all possible } x(u),\ 0 \leq u < s\end{aligned}$$

In other words, when the *present* state of the process is known, the future development of the process is independent of anything that has happened in the past.

Markov process

A 'continuous time' stochastic process that fulfills the Markov property is called a *Markov process.*

We will further assume that the Markov process for all i, j in $\mathcal{X}$ fulfills

$$\Pr(X(t+s)=j \mid X(s)=i) = \Pr(X(t)=j \mid X(0)=i) \quad \text{for all } s,t \geq 0$$

which says that the probability of a transition from state i to state j does not depend on the global time and only depends on the time interval available for the transition.

A process with this property is known as a process with *stationary transition probabilities*, or as a *time-homogeneous* process.

Transition probabilities

The transition probabilities of the Markov process

$$P_{ij}(t) = \Pr(X(t) = j \mid X(0) = i)$$

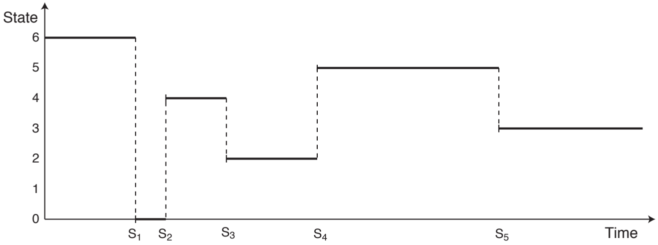
may be arranged as a matrix

$$\mathbb{P}(t) = \begin{pmatrix} P_{00}(t) & P_{01}(t) & \cdots & P_{0r}(t) \\ P_{10}(t) & P_{11}(t) & \cdots & P_{1r}(t) \\ \vdots & \vdots & \ddots & \vdots \\ P_{r0}(t) & P_{r1}(t) & \cdots & P_{rr}(t) \end{pmatrix}$$

When a process is in state i at time 0, it must either be in state i at time t or have made a transition to a different state. We must therefore have

$$\sum_{j=0}^{r} P_{ij}(t) = 1$$

Trajectory



Let $0 = S_0 \leq S_1 \leq S_2 \leq \cdots$ be the times at which transitions occur, and let $T_i = S_{i+1} - S_i$ be the ith interoccurrence time, or *sojourn time*, for $i = 1, 2, \ldots$.

A possible trajectory of a Markov process is illustrated above.

We define S_i such that transition i takes place immediately before S_i, in which case the trajectory of the process is continuous from the right.

Time in state

A Markov process enters state i at time 0, such that $X(0) = i$. Let $\widetilde{T}_i$ be the sojourn time in state i. [Note that T_i denotes the ith interoccurrence time, while $\widetilde{T}_i$ is the time spent during a visit to state i.]

We want to find $\Pr(\widetilde{T}_i > t)$. We observe that the process is still in state i at time s, that is, $\widetilde{T}_i > s$, and are interested in finding the probability that it will remain in state i for t time units more. We hence want to find $\Pr(\widetilde{T}_i > t + s \mid \widetilde{T}_i > s)$.

Since the process has the Markov property, the probability for the process to stay for t more time units is determined only by the current state i. The fact that the process has been staying there for s time units is therefore irrelevant. Thus

$$\Pr(\widetilde{T}_i > t + s \mid \widetilde{T}_i > s) = \Pr(\widetilde{T}_i > t) \quad \text{for } s, t \geq 0$$

Hence $\widetilde{T}_i$ is *memoryless* and must be exponentially distributed.

Sojourn times

From the previous frame we can conclude that the sojourn times $T_1, T_2, \ldots$ are independent and exponentially distributed. The independence follows from the Markov property.

Skeleton

We will now disregard the time spent in the various states and only consider the transitions that take place at times $S_1, S_2, \ldots$.

Let $X_n = X(S_n)$ denote the state immediately after transition n. The process $\{X_n, n = 1, 2, \ldots\}$ is called the *skeleton* of the Markov process.

Transitions of the skeleton may be considered to take place at discrete times $n = 1, 2, \ldots$. The skeleton may be imagined as a chain where all the sojourn times are deterministic and of equal length. It is straightforward to show that the skeleton of a Markov process is a discrete-time Markov chain; see Ross (1996). The skeleton is also called the *embedded* Markov chain.

Alternative construction

We may construct a Markov process as a stochastic process having the properties that each time it enters a state i:

1. The amount of time $\widetilde{T}_i$ the process spends in state i before making a transition into a different state is exponentially distributed with rate, say α_i.
2. When the process leaves state i, it will next enter state j with some probability P_{ij}, where $\sum_{\substack{j=0 \\ j \neq i}}^{r} P_{ij} = 1$.

The mean sojourn time in state i is therefore

$$E(\widetilde{T}_i) = \frac{1}{\alpha_i}$$

Transition rates

Let a_{ij} be defined by

$$a_{ij} = \alpha_i \cdot P_{ij} \quad \text{for all } i \neq j$$

Since α_i is the rate at which the process leaves state i and P_{ij} is the probability that it goes to state j, it follows that a_{ij} is the rate when in state i that the process makes a transition into state j. We call a_{ij} the *transition rate* from i to j.

Since $\sum_{j \neq i} P_{ij} = 1$, it that

$$\alpha_i = \sum_{\substack{j=0 \\ j \neq i}}^{r} a_{ij}$$

Transition rates (2)

Let T_{ij} be the time the process spends in state i before entering into state $j(\neq i)$. The time T_{ij} is exponentially distributed with rate a_{ij}.

Consider a short time interval Δt. Since T_{ij} and $\widetilde{T}_i$ are exponentially distributed, we have that

$$P_{ii}(\Delta t) = \Pr(\widetilde{T}_i > \Delta t) = e^{-\alpha_i \Delta t} \approx 1 - \alpha_i \Delta t$$
$$P_{ij}(\Delta t) = \Pr(T_{ij} \leq \Delta t) = 1 - e^{-a_{ij} \Delta t} \approx a_{ij} \Delta t$$

when Δt is “small”. We therefore have that

$$\lim_{\Delta t \to 0} \frac{1 - P_{ii}(\Delta t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Pr(\widetilde{T}_i < \Delta t)}{\Delta t} = \alpha_i$$
$$\lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Pr(T_{ij} < \Delta t)}{\Delta t} = a_{ij} \quad \text{for } i \neq j$$

Transition rate matrix

We can deduce α_i and P_{ij} when we know a_{ij} for all i, j, and may therefore define the Markov process by specifying (i) the state space $\mathcal{X}$ and (ii) the transition rates a_{ij} for all $i \neq j$ in $\mathcal{X}$. The second definition is often more natural and will be our main approach in the following.

We may arrange the transition rates a_{ij} as a matrix:

$$\mathbb{A} = \begin{pmatrix} a_{00} & a_{01} & \cdots & a_{0r} \\ a_{10} & a_{11} & \cdots & a_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r0} & a_{r1} & \cdots & a_{rr} \end{pmatrix}$$

where the diagonal elements are

$$a_{ii} = -\alpha_i = -\sum_{\substack{j=0 \\ j \neq i}}^{r} a_{ij}$$

Departure rates

Observe that the entries of row i of the transition rate matrix $\mathbb{A}$ are the transition rates out of state i (for $j \neq i$). We will call them *departure rates* from state i.

The rate $-a_{ii} = \alpha_i$ is the *total* departure rate from state i.

The sum of the entries in row i is always equal to 0, for all $i \in \mathcal{X}$.

Chapman-Kolmogorov equations

By using the Markov property and the law of total probability, we realize that

$$P_{ij}(t+s) = \sum_{k=0}^{r} P_{ik}(t)P_{kj}(s) \quad \text{for all} \quad i,j \in \mathcal{X}, t,s > 0$$

These equations are known as the *Chapman-Kolmogorov equations.* The equations may be written in matrix terms as

$$\mathbb{P}(t+s) = \mathbb{P}(t) \cdot \mathbb{P}(s)$$

Notice that $\mathbb{P}(0) = \mathbb{I}$ is the identity matrix, and that when t is an integer, we have that $\mathbb{P}(t) = [\mathbb{P}(1)]^t$. It can be shown that this also holds when t is not an integer.

Kolmogorov differential equations

To find $P_{ij}(t)$ we start by considering the Chapman-Kolmogorov equations

$$P_{ij}(t+\Delta t) = \sum_{k=0}^{r} P_{ik}(t) P_{kj}(\Delta t)$$

We consider

$$P_{ij}(t+\Delta t) - P_{ij}(t) = \sum_{\substack{k=0 \\ k\neq j}}^{r} P_{ik}(t) P_{kj}(\Delta t) - [1 - P_{jj}(\Delta t)] P_{ij}(t)$$

By dividing by Δt and then taking the limit as $\Delta t \to 0$, we obtain

$$\lim_{\Delta t \to 0} \frac{P_{ij}(t+\Delta t) - P_{ij}(t)}{\Delta t} = \lim_{\Delta t \to 0} \left[\sum_{\substack{k=0 \\ k\neq j}}^{r} P_{ik}(t) \frac{P_{kj}(\Delta t)}{\Delta t} - \frac{1 - P_{jj}(\Delta t)}{\Delta t} P_{ij}(t) \right]$$

Kolmogorov differential equations (2)

Since the summing index is finite, we may interchange the limit and summation and obtain

$$\dot{P}_{ij}(t) = \sum_{\substack{k=0 \\ k \neq j}}^{r} a_{kj} P_{ik}(t) - \alpha_j P_{ij}(t) = \sum_{k=0}^{r} a_{kj} P_{ik}(t)$$

where, as before, $a_{jj} = -\alpha_j$. The differential equations above are known as the Kolmogorov *forward equations.*

Kolmogorov differential equations (3)

The Kolmogorov forward equations may be written in matrix format as

$$\mathbb{P}(t) \cdot \mathbb{A} = \dot{\mathbb{P}}(t) \tag{1}$$

State equations – 1

Assume that the Markov process is in state i at time 0, that is, $X(0) = i$. This means that

$$P_i(0) = \Pr(X(0) = i) = 1$$
$$P_k(0) = \Pr(X(0) = k) = 0 \quad \text{for } k \neq i$$

Since we know the state at time 0, we may simplify the notation by writing $P_{ij}(t)$ as $P_j(t)$.

The vector $\mathbf{P}(t) = [P_0(t), P_1(t), \ldots, P_r(t)]$ then denotes the distribution of the Markov process at time t, when we *know* that the process started in state i at time 0. We further know that $\sum_{j=1}^{r} P_j(t) = 1$.

State equations – 2

The distribution $\mathbf{P}(t)$ may be found from the Kolmogorov forward equations

$$\dot{P}_j(t) = \sum_{k=0}^{r} a_{kj} P_k(t)$$

where $a_{jj} = -\alpha_j$. In matrix terms, this may be written

$$[P_0(t), \ldots, P_r(t)] \cdot \begin{pmatrix} a_{00} & a_{01} & \cdots & a_{0r} \\ a_{10} & a_{11} & \cdots & a_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r0} & a_{r1} & \cdots & a_{rr} \end{pmatrix} = \left[\dot{P}_0(t), \ldots, \dot{P}_r(t)\right]$$

or in a more compact form as

$$\mathbf{P}(t) \cdot \mathbb{A} = \dot{\mathbf{P}}(t)$$

These equations are called the *state equations* for the Markov process

Example 8.5 – 1

Consider a single component with states:

1 The component is functioning
0 The component is in a failed state

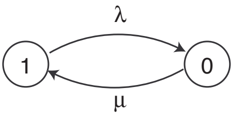
The transition rates are:

$a_{10} = \lambda$ The failure rate of the component
$a_{01} = \mu$ The repair rate of the component

The mean sojourn time in state 1 is MTTF $= 1/\lambda$, and the mean sojourn time in state 0 is the mean downtime, MDT $= 1/\mu$. The mean downtime is sometimes called the mean time to repair (MTTR).

Example 8.5 – 2)

The state transition diagram for the single component is



The state equations are

$$[P_0(t), P_1(t)] \cdot \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix} = \left[\dot{P}_0(t), \dot{P}_1(t)\right]$$

The component is assumed to be functioning at time $t = 0$,

$$P_1(0) = 1, \qquad P_0(0) = 0$$

and $P_0(t) + P_1(t) = 1$ for all t.

Example 8.5 – 3

The solution is

$$P_1(t) = \frac{\mu}{\mu+\lambda} + \frac{\lambda}{\mu+\lambda}\, e^{-(\lambda+\mu)t}$$

$$P_0(t) = \frac{\lambda}{\mu+\lambda} - \frac{\lambda}{\mu+\lambda}\, e^{-(\lambda+\mu)t}$$

The limiting availability is

$$P_1 = \lim_{t\to\infty} P_1(t) = \frac{\mu}{\lambda+\mu}$$