PK8207 - Lecture memo

Jørn Vatn
Email: jorn.vatn@ntnu.no

Updated 2022-03-01

Markov State Model - An introduction

Introduction

Consider a stochastic process $\{Y(t), t \in \Theta\}$, where $Y(t)$ describes the state (deterioration level) of an item at time t. In the following we will assume that the state variable only takes a finite number of states. We first present the model when no maintenance is carried out, i.e., we start at time $t = 0$ and observe the system until failure. Let:

$$\begin{aligned} Y(0) &= y_0 \\ Y(T) &= y_r \end{aligned} \tag{1}$$

where T per definition is the time of the first failure. Between y_0 and y_r there are $r-1$ intermediate sates. By choosing a large value of r we could obtain a very good approximation to a continuous process if this is required. We will now let $\tilde{T}_i, i = 0, \ldots, r-1$ be sojourn times, i.e., how long the system stay in state i. Notationally we will typically denote the states by their number rather than by the value to simplify notation.

For the initial model we assume that the sojourn times are independent and exponentially distributed with parameter λ_i. Later on we will investigate how sojourn times may be modelled by arbitrary distributions. We also assume that the process runs through all states chronologically from y_0 to y_r without "stepping back" at any time.

Before we present the modelling framework for this simple situation we introduce the maintenance model. Figure 1 depicts the development of $Y(t)$ as a function of time. On the x-axis it is indicated that the system is inspected at period of times $\tau, 2\tau, 3\tau, \ldots$. If the system is found in state $Y(t) \geq y_l$ at an inspection, the system is renewed to an as good as new state, i.e., y_0.

We now go back to the simple situation where maintenance is not considered. Let $P_i(t)$ denote the probability that the system is in state i at time t. By standard Markov considerations we obtain the Markov differential equations:

$$P_i(t + \Delta t) \approx P_i(t)(1 - \lambda_i \Delta t) + P_{i-1}(t)\lambda_{i-1}\Delta t \tag{2}$$

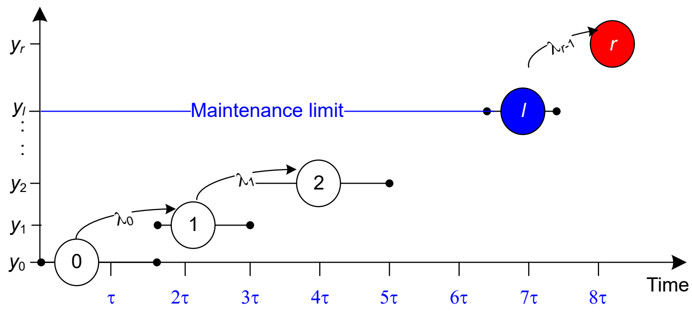


Figure 1: Markov transition diagram

where Δt is a small time interval and and we set $\lambda_{-1} = 0$ per definition. Further the initial conditions are given by:

$$P_0(0) = 1$$
$$P_i(0) = 0 \text{ for } i > 0 \tag{3}$$

Equation (2) could easily be integrated by a computer program, for example VBA in MS Excel. It is now easy to find MTTF by another integration, i.e.,

$$\text{MTTF} = \int_0^\infty R(t)dt = \int_0^\infty [1 - P_r(t)]\,dt \tag{4}$$

and we should verify that we get $\text{MTTF} = \sum_{i=0}^{r-1} \lambda_i^{-1}$. Note that the transition rates, λ_i's, are assumed to be known, that is either they are estimated from data, or found by expert judgement exercises.

Exercise 1

Assume $r = 5$ and $\lambda_i = 0.01, i = 0, 1, \ldots$. Integrate the Markov differential equations and obtain the expected value and variance of the time to failure. Hint: Use partial integration for the variance similar to $\text{MTTF} = \int R(t)dt$. □

Equation (2) may be used in situations where we only allow transitions from state i to state $i+1$. In more general situations there could be transitions in principle from any state i to any state j. In this situation we need to work with matrices. Let $\mathbf{A}$ be an $(r+1) \times (r+1)$ matrix where element (i, j) represents the constant transition rate from state i to state j. The indexing here starts at 0, e.g., $\mathbf{A}(0,1) = a_{0,1}$ is the transition from state 0 to state 1.

Further, let $\mathbf{P}(t)$ be the time dependent probability vector for the various states defined in $\mathbf{A}$. We now let $\mathbf{P}(t = 0) = [1, 0, 0, \ldots, 0]$ to reflect that the system starts in state 0. From standard Markov theory we now need the Markov differential equations, i.e., $\mathbf{P}(t) \cdot \mathbf{A} = \dot{\mathbf{P}}(t)$, from which it follows:

$$\mathbf{P}(t + \Delta t) \approx \mathbf{P}(t)[\mathbf{A}\Delta t + \mathbf{I}] \tag{5}$$

where Δt is a small time interval. Equation (5) is now used repeatedly to find the time dependent solution for the entire system. This corresponds to integrating Equation (2).

We now outline the main principle for working with matrices to find the time dependent solution and other relevant quantities. Assume we have access to a small library of matrix routines:

```
Function mMult(A,B) -> Returns a matrix equal to  A * B
Subroutine fixA(A) -> Fill diagonal of A such that sumrow=0
Function getIntMatrix(A, DeltaT) -> [A * DeltaT + I]
```

In the following we assume that the matrix library is defined by standard indexing, i.e., the first row is denoted row number 1 and so on. A warm up exercise to find MTTF is now:

```
Function getMTTF(A)
fixA A
MTTF = initial guess
DeltaT = MTTF / 1000
hlp = 0
t=0
P=[1,0,0,....]
IM = getIntMatrix(A, DeltaT)
Do While t < 5*MTTF
   P = mMult(P, IM)
   hlp = hlp +  (1-P(r+1)) * DeltaT
   t = t + DeltaT
Loop
getMTTF = hlp
End Function
```

To get higher precision we could increase the integration to e.g., 10MTTF. Note the motivation for this approach is given by:

$$\text{MTTF} = \int_0^{\infty} R(t)dt = \int_0^{\infty} [1 - P_r(t)]\,dt \tag{6}$$

where $1 - P_r(t)$ is the probability that we are not in state r at time t.

So far the maintenance regime is not reflected in the approach. Let $\lambda_{\text{E}}(\tau, l)$ be the effective failure rate, i.e., the expected number of failures per unit time if the system is inspected every τ time unit, and renewed whenever $Y(t) \geq y_l$ at an inspection. In the integration of Equation (5) we start with $t = 0$ and whenever t coincides with τ, 2τ etc., special actions are taken:

```
Function lambdaEffective(A,tau,l)
fixA A
```

```
MTTF = getMTTF(A)
DeltaT = MTTF / 1000
hlpF = 0
t=0
localTime=0
P=[1,0,0,....]
IM = getIntMatrix(A, DeltaT)
Do While t < 10*MTTF
   P = mMult(P, IM)
   hlpF = hlpF + P(r + 1)      Add to effective failure rate
   P(1) = P(1) + P(r + 1)      If system is failed, it is assumed to be renewed
   P(r + 1) = 0                Clear probability
   If localTime >= tau Then
     sumP = 0
     For i = l+1 To r
         SumP = SumP + P(i)
         P(i)=0
      Next i
      P(1) = P(1) + SumP
      localTime = 0
   Else
      localTime = localTime + DeltaT
   End If
   t = t + DeltaT
Loop
lambdaEffective = hlpF / t
End Function
```

Note the indexing, i.e., the failed state is $r+1$ and the maintenance limit is $l+1$.

In the `If localTime = tau` part of the script above we have used a loop to simulate what is happening during an inspection. A more efficient way to do this would be to create an "inspection matrix", say $\mathbf{M}$ defined by:

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{7}$$

where the starting point is an identity matrix, but where we from the row corresponding to state l shift the "ones" to the left.

```
:
If localTime >= tau Then
   P = mMult(P, M)
   localTime = 0
Else
:
```

Such an inspection matrix could also be used to specify that an inspection is not perfect. For example if q is the probability that an inspection fails to reveal that the actual state is l or higher, the corresponding leftmost "one" is replaced by $1-q$ and the diagonal element is replaced by q for rows corresponding to states $l, l+1, \ldots, r-1$. A inspection matrix could also be used to specify that upon an inspection it might be decided to repair to a state which is not as good as new. For example in 80% of the cases we repair to state 0, in 15% of the cases we repair to sate 1 and in 5% of the cases we repair to state 2.

Exercise 2

Assume $r = 5$ and $\lambda_i = 0.01, i = 0, 1, \ldots$ (time unit weeks). Assume the system is inspected with intervals of length $\tau = 26$. If the system is found in state $r = 4$ the system will be renewed. Renewal takes place immediately. The probability that a inspection reveals that the system is in state $r = 4$ is 70% when this is the case. Find the effective failure rate for this situation. □

Significant repair times

So far we have assumed that repair times could be neglected. If we can not neglect repair times we need to model repair times in the transition matrix **A**. For example if at an inspection we with some probability q will decide to repair from state i to state j with constant repair rate μ a first approach would be to modify the **A**-matrix, i.e., $\mathbf{A}(i,j) = a_{i,j} = q\mu$. However, this would imply that a repair starts immediately after the system has reached state j. In reality, a repair can first start after the coming inspection.

To handle the situation we now introduce "virtual" states. A virtual state is a state in the **A**-matrix representing the situation where a maintenance action has been decided and the repair is actually started. For each pair (i,j) where there could be a repair from physical state i to physical state j a virtual state $k_{i,j}$ is defined. Then the associated transition rate is set to $a_{k_{i,j},j} = \mu$. The row and column representing the virtual state $k_{i,j}$ can be any ones larger than those already "occupied". The inspection matrix **M** will also get an additional row and column representing the virtual state $k_{i,j}$, where $\mathbf{M}(i, k_{i,j}) = q$, where we in addition need to ensure that the row sum equals one.

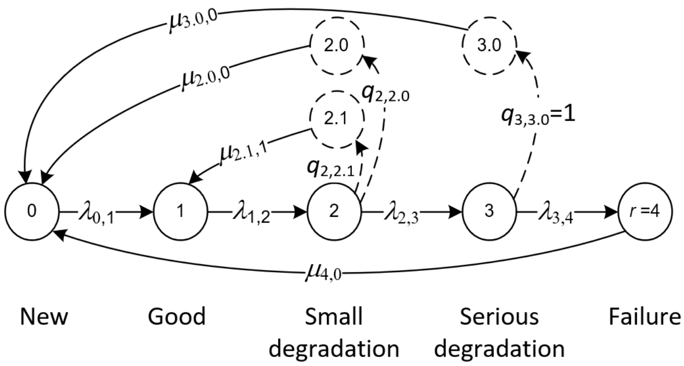


Figure 2: Markov transition diagram with potential repairs

Note that while repairing from state i to state j represented by $a_{k_{i,j},j} = \mu$ there might be a "competing" transition from for example state i to state l, thus we also need to specify $a_{k_{i,j},l} = \lambda_{i,l}$. Such transitions are not shown in Figure 2.

Figure 2 illustrates the Markov diagram for a situation with $r = 4$. Here λ_{i_j} is the transition rate from state i to state j representing degradation. Further $\mu_{k_{i,j},j}$ is the repair rate from virtual state $k_{i,j}$ to state j. When a repair is initiated as a result of a proof-test, virtual states are introduced. For example the state (2.1) represent that after a test it is decided to repair from state 2 to state 1. The doted lines represent transitions that instantaneously take place after a proof-test. The probabilities given by the q-values represent maintenance decisions. For example $q_{3,3.0} = 1$ represents that if a state 3 is revealed by a proof-test, we always initiate a repair to state 0. $q_{2,2.0}$ is representing the probability that we after revealing a state 2 on a proof-test we initiate a repair to state 0. The q-values are entered into the inspection matrix, **M**.

In Figure 2 there are three nodes representing that the system is in a "small degradation" state, i.e., physical state 2. State 2 in the diagram is a *hidden state*, we are not aware of any transition from state 1 to state 2. The states (2.0) and (2.1) are *evident states*, we know that we are in main state 2 (small degradation), a maintenance request has been issues (to state 0 and state 1 respectively).

Note that in Figure 2 we use the notation $a_{\text{From},\text{To}}$ without indicating the actual row and column *numbers* in the transition matrix. The notation $a_{k_{i,j},l}$ on the other hand, is used to identify a row and column number in a matrix in the code when we do the programming.

In previous sections we have focused on the effective failure rate, but we might also be interested in the average portion of time we are in each state. For example we may use:

:

```
Do While t < 10*MTTF
   P = mMult(P, IM)
   Pavg = Pavg + P
   If localTime >= tau Then
      P = mMult(P, M)
      localTime = 0
   Else
      localTime = localTime + DeltaT
   End If
   t = t + DeltaT
Loop
Pavg = Pavg * DeltaT / t
:
```

Exercise 3

Assume $r = 5$ and $\lambda_i = 0.01, i = 0, 1, \ldots$ (time unit weeks). Assume the system is inspected with intervals of length $\tau = 26$. If the system is found in state $r = 4$ the system will be renewed. There is a *logistic delay* of in average 4 weeks before the repair takes place. Delay time is assumed to be exponentially distributed. The probability of revealing state $r = 4$ is still 70%. Find the effective failure rate for this situation. □

Least Square Estimation - Covariates

In this section we present a least square approach for parameter estimation. The main idea is to compare the actual transitions taking place between observations with the expected transitions given the parameters in the model. The approach allows for also treating systems operated under various loads, i.e., with different values of explanatory variables (covariates).

Assume that the system is in state s at time t and we consider a later point of time $t + u$ where no maintenance has been conducted in the period between. We have

$$\begin{aligned} P_s(t) &= 1 \\ P_i(t) &= 0 \text{ for } i \neq s \end{aligned} \tag{8}$$

The expected system state at time $t + u$ is then given by:

$$\mathrm{E}(Y(t+u)|Y(t) = s) = \sum_{i=s}^{r} i P_i(t+u) \tag{9}$$

Note that we sum over states larger or equal to s since we assume that the system cannot improve during the period from t to $t + u$. In the more general situation where we also would like to estimate repair rates we need to sum over all possible states.

For a given set of parameters in the Markov transition matrix we may find the expected value by integrating the Markov equations u time units from t and then use Equation (9). For the integration, i.e., finding the $P_i(t+u)$'s we need to repeatedly use Equation (5). In our LS approach later on we need numerical methods implying that the numerical integration of the Markov equations will be repeated very many times for different parameter values. To save computational time we propose a more efficient approach. A similar approach is in fact used in many algorithms for calculating the exponential of matrices. We have:

$$\mathbf{P}(t+u) \approx \mathbf{P}(t)[\mathbf{A}\Delta t + \mathbf{I}]^{2^n} \tag{10}$$

where $\Delta t = u/2^n$ and n is sufficient large to get a low value of Δt. Typically $n = 10$. To raise $[\mathbf{A}\Delta t + \mathbf{I}]$ to the power 2^n we first calculate $\mathbf{E} = [\mathbf{A}\Delta t + \mathbf{I}]$, and then we repeat n times: $\mathbf{E} = \mathbf{E}^2$.

Obvious, if we have a library function that takes exponential of matrices, we may directly use $\mathbf{P}(t+u) = \mathbf{P}(t)e^{\mathbf{A}u}$ as an alternative.

Also note that for our data which typically is sampled with fixed intervals, say τ, we only need to calculate $[\mathbf{A}\tau/2^n + \mathbf{I}]^{2^n}$ once for all observation for a given combination of the parameter vector to speed up the process. In other situations we have data sets where the observation period varies, then it is convenient to save *several* “multiplication matrices", i.e., $[\mathbf{A}\tau_i/2^n + \mathbf{I}]^{2^n}$.

To estimate parameters in the Markov State model we assume the following data is available:

- Several systems are observed, each system has one or more observations
- An observation, say j, from one system comprise
  - The state, s_j, of the system at time t_j and the state, s_{j^-}, of that system at the previous time, t_{j^-}, the system were observed
  - A vector of explanatory variables (stressors) $\mathbf{z}_j = [z_1, z_2, \ldots]$ for the time interval between t_j and t_{j^-}
- We let j be an index running through the all observations, i.e., $j = 1, 2, \ldots, J$ for all systems

We assume that each transition rate in the Markov matrix could be written on the form

$$\lambda = e^{\beta_0 + \beta_1 z_1 + \beta_2 z_2 \ldots} \tag{11}$$

A least square approach to estimate $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \ldots]$ is now given by minimizing Equation (12):

$$Q(\boldsymbol{\beta}) = \sum_{j=1}^{J} \left[s_j - \mathrm{E}(Y_j(t_j) | Y_{j^-}(t_{j^-}) = s_{j^-}) \right]^2 \tag{12}$$

where we need Equation (9) to find $E(Y_j(t_j)|Y_{j^-}(t_{j^-}) = s_{j^-})$.

The procedure simplifies if we assume $\beta_1, \beta_2, \ldots$ are the same for all transition rates, and that only β_0 varies between the various transition rates.

Exercise 4

Assume we have 4 systems each with with states $1, 2, \ldots, r$ where $\lambda_{i+1} = (1 + v)\lambda_{i+1}$. The systems are proof-tested every τ time units.

Numerical values are given by $r = 5, \lambda_1 = 0.0001, v = 0.2$, and $\tau = 730$. The maintenance limit is $l = r - 1$. Upon a proof test nothing is done for states $1, 2, \ldots, l - 1$. If a system is in a state $\geq l$ at a proof-test, an instantaneous repair takes place bringing the system back to state 1.

Use Monte Carlo simulation to simulate the observation set. Assume you simulate over 60 months (5 years).

Exercise 5

In this exercise we will use the data from the previous exercise. There are only two parameters to estimate, i.e., λ_1 and v. To get an initial guess for λ_1 we may do the following:

1. Count the number of situations in the data set where there is a transition from state 1 to another state, and let this number be denoted n_1
2. Count the number of occurrences where one system remains in state 1 from one inspection to the next inspection, or jumps from state 1 to another state from one inspection to the next inspection. Let this number be denoted t_1
3. An initial guess for λ_1 is now given by $\hat{\lambda}_1 = n_1/t_1$

We can repeat for $\lambda_2, \lambda_3, \ldots, \lambda_{l-1}$. Note that we cannot estimate the transition rate into the fault state, i.e., λ_l by this procedure because there are no observed jumps from state l to state r due to our maintenance strategy.

We may now get an initial guess for v by $\hat{v} = \hat{\lambda}_2/\hat{\lambda}_1 - 1$. We could also use $\hat{v} = \hat{\lambda}_3/\hat{\lambda}_2 - 1$ and so forth, so an average of these v-values could be a reasonable approach to obtain an initial estimate, $\hat{v}$.

a) Calculate initial estimates for λ_1 and v as indicated above for your simulated dataset

b) Keep v fixed, find the LS-estimate for λ_1 according to the procedure described. Use any numerical routine for minimizing a univariate function.

c) Keep λ_1 fixed, i.e., the estimate from b), and find the LS-estimate for v

d) Keep v fixed, i.e., from c), and find the LS-estimate for λ_1 according to the procedure described.

e) Keep λ_1 fixed, i.e., the estimate from d), and find the LS-estimate for v

f) Compare the result with using a minimization routine that allows for several variables.

Phase type modelling

So far we have assumed that the sojourn times are exponentially distributed. This assumption could be questioned if there are failure mechanisms like wear, fatigue, corrosion etc. that drives the degradation of the system. Phase type modelling is an approach where we may approximate a stochastic variable with a multi state Markov model. The more states we use the better will the approximation be. In the following we do not discuss the explicit fitting of model parameters for the approximation. Several statistical packages exist for this. We will also assume that each random variable is approximated by a two-state Markov model in order to reduce the total number of states.

Model

Consider a system having three main states, 1 = new, 2 = degraded, and F = failed. With *main state* we here mean what the categorization used by the maintenance department.

We assume that sojourn times in state 1 and state 2 could be approximated by a phase type distribution, i.e., $\tilde{T}_1 \sim F_1(t)$ and $\tilde{T}_2 \sim F_2(t)$. The sojourn times are assumed to be stochastically independent.

For $\tilde{T}_1$ the phase type model comprises two sub states, i.e., $1a$ and $1b$. A acyclic phase type model is used where:

- The probability that the system starts in sub state $1a$ is p_1, and the probability that the system starts in sub state $1b$ is $1-p_1$
- There is a constant transition rate, λ_{1a} from state $1a$ to $1b$
- There is a constant transition rate, λ_{1b} from state $1b$ to an absorbing state outside the system

A similar model exist for sojourn time 2.

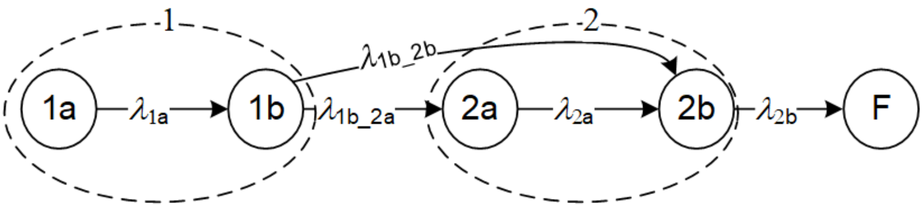
Note that given $F_1(t)$ and $F_2(t)$ we may in principle find the distribution of the time to failure for this system by the convolution theorem. This will not be pursued here.

System modelling

The phase type models for the two sojourn time enable easy integration by use of the Markov equations $\mathbf{P}(t+\Delta t) \approx \mathbf{P}(t)[\mathbf{A}\Delta t + \mathbf{I}]$ when each sojourn time is treated independently.

In order to have a complete model we need to link the two phase type models.

Proposition: For the phase type distribution approximating the first sojourn time there is a rate λ_{1b} from state $1b$ to an absorbing state outside the system. This transition is split into two transitions, one to state $2a$ and one to state $2b$. The corresponding rates are $\lambda_{1b\_2a} = p_2\lambda_{1b}$ and $\lambda_{1b\_2b} = (1-p_2)\lambda_{1b}$ respectively. The situation is shown in the figure below:



The dashed ellipses represent the physical or main states "new" and "degraded" as observed by, e.g., maintenance personnel. The states $1a$, $1b$, $2a$ and $2b$ are artificial or sub states used for modelling, but have no physical interpretation.

Let $P_F(t)$ be the probability that the system is in state F at time t. $P_F(t)$ can easily be obtained by integration the Markov equation for the compound system yielding the cumulative distribution function for time to system failure.

Proof

The idea behind the proof is to first find the cumulative distribution function for the time to system failure without making any assumption regarding how to link the two phase type distribution. Then we repeat the procedure when the two phase type distributions are linked in the proposed manner. If this gives the same result, our proposition is valid.

First define $\tilde{T}_{2a}$ and $\tilde{T}_{2b}$ to be the sojourn times for states $2a$ and $2b$ respectively. It follows that

$$\Pr(\tilde{T}_2 \le t) = p_2 \Pr(\tilde{T}_{2a} + \tilde{T}_{2b} \le t) + (1-p_2)\Pr(\tilde{T}_{2b} \le t) \tag{13}$$

Let T be the time to failure. Now assume that the first sojourn time $\tilde{T}_1 = x$. It follows that:

$$\Pr(T = \tilde{T}_2 + x \le t | \tilde{T}_1 = x) = p_2 \Pr(\tilde{T}_{2a} + \tilde{T}_{2b} + x \le t) + (1-p_2)\Pr(\tilde{T}_{2b} + x \le t) \tag{14}$$

Equation (14) does not make any assumption regarding how the two phase type models are linked. It just prescribes a way to obtain the cumulative distribution function (CDF) for the system failure time, given that the first sojourn time is x. To obtain the unconditional CDF we integrate over all possible x-values, i.e.,

$$\Pr(T \leq t) = \int_0^\infty \Pr(\tilde{T}_2 \leq t - x) P_{1b}(x) \lambda_{1b} dx \tag{15}$$

where $P_{1b}(x)$ is the time dependent probability that the first phase type system is in state $1b$ at time x. Using equation (14) we get:

$$\Pr(T \leq t) = \int_0^\infty p_2 \Pr(\tilde{T}_{2a} + \tilde{T}_{2b} \leq t - x) P_{1b}(x) \lambda_{1b} dx + \int_0^\infty (1 - p_2) \Pr(\tilde{T}_{2b} \leq t - x) P_{1b}(x) \lambda_{1b} dx \tag{16}$$

Now, consider the compound model shown in the figure. In the compound model there are to ways we may leave state $1b$, i.e., there is a transition from the “physical” state 1 to the “physical” state 2. These to possibilities are either going from state $1b$ to state $2a$ or going from state $1b$ to state $2b$.

Assume there is a transition from state $1b$ to state $2a$ at time x. Let this event be denoted A_x, giving the conditional CDF:

$$\Pr(T \leq t | A_x) = \Pr(\tilde{T}_{2a} + \tilde{T}_{2b} \leq t - x) \tag{17}$$

Similarly, let B_x be the event that there is a transition from state $1b$ to state $2b$ at time x, giving the conditional CDF:

$$\Pr(T \leq t | B_x) = \Pr(\tilde{T}_{2b} \leq t - x) \tag{18}$$

Since A_x and B_x are disjoint we integrate over all possible values to get the unconditional CDF:

$$\Pr(T \leq t) = \int_0^\infty \Pr(\tilde{T}_{2a} + \tilde{T}_{2b} \leq t - x) P_{1b}(x) \lambda_{1b\_2a} dx + \int_0^\infty \Pr(\tilde{T}_{2b} \leq t - x) P_{1b}(x) \lambda_{1b\_2b} dx \tag{19}$$

which is the same as equation (16) provided $\lambda_{1b\_2a} = p_2 \lambda_{1b}$ and $\lambda_{1b\_2b} = (1 - p_2) \lambda_{1b}$.

This means that we have proven the following:

- Given that we approximate two independent sojourn times with two phase type distributions having the structure presented above:
- The two phase type models can be combined as illustrated in the figure
- Yielding the correct CDF for the sum of these two sojourn times which has been the objective to prove

It should be rather straight forward to prove this for the general case with more than two "physical" states.

For the model with the artificial states, the Markov property holds. However, if we consider the three main states (1=new, 2=degraded, and F=failed), the Markov property does not hold. Given that we have stayed in main state 2 for some time, the probability of escaping that state will typically increase as time goes by. The explanation is that given we entered state 2 at time x, we either went to the artificial state $2a$ with probability p or to state $2b$ with probability $1-p$, which is exactly matching the phase type model we use for the "physical" state 2 when the "failure rate function" for state 2 is increasing.

Varying intervals for inspection

The idea of having fixed lengths between inspections may be questioned. Obvious it would be some administrative advantages if we could stick to the same inspection intervals independent on the system state. On the other hand it seems reasonable to reduce the inspection intervals as we approach the maintenance limit. To obtain the total failure rate, number of inspections, and number of repairs for a general inspection and maintenance strategy is almost impossible. In the following we present an approach where we make some assumptions which would be reasonable to handle from an administrative point of view, and which also should not be far from the optimal solution:

- The time intervals between inspections are either τ_{L} (long), τ_{M} (medium), or τ_{S} (short). Further k_{L} and k_{M} are integers such that $\tau_{\mathrm{L}} = k_{\mathrm{L}}\tau_{\mathrm{S}}$ and $\tau_{\mathrm{M}} = k_{\mathrm{M}}\tau_{\mathrm{S}}$.
- $\mathscr{L}_{\mathrm{L}}$ is a set of states which require inspection interval τ_{L}, $\mathscr{L}_{\mathrm{M}}$ is a set of states which require inspection interval τ_{M} and $\mathscr{L}_{\mathrm{S}}$ is a set of states which require inspection interval τ_{S}. For all other states, $\mathscr{L}_{\mathrm{R}}$, it is required to repair the system to a good as new state.
- If a failure occurs at time t in between inspections the system will be repaired to a good as new condition immediately, and the first inspection of length τ_{L} will take place at the smallest value of $k\tau_{\mathrm{S}}$ where k is an integer and $k\tau_{\mathrm{S}} > t + \tau_{\mathrm{L}}$.

In the modelling we assume that we at any time, $t \neq k\tau_{\mathrm{S}}$, have:

- The current inspection regime is governed by τ_{L}, τ_{M} or τ_{S}
- Let t_{C} be the starting point of the current inspection interval, i.e., $t_{\mathrm{C}} > k\tau_{\mathrm{S}}$ and $t_{\mathrm{C}} < (k+1)\tau_{\mathrm{S}}$ for some integer k
- If the current inspection regime is governed by τ_{L}, the next inspection will take place at one of the following point of times: $t_{\mathrm{C}} + \tau_{\mathrm{S}}, t_{\mathrm{C}} + 2\tau_{\mathrm{S}}, \ldots, t_{\mathrm{C}} + k_{\mathrm{L}}\tau_{\mathrm{S}}$

- If the current inspection regime is governed by τ_M, the next inspection will take place at one of the following point of times: $t_C + \tau_S, t_C + 2\tau_S, \ldots, t_C + k_M\tau_S$

We may now define different **P**-vectors to hold the time dependent state probabilities as we integrate the solution. Let $\mathbf{P}_S(t)$ be defined such that

$$P_{i,S}(t) = \Pr(Y(t) = y_i \cap \text{ current regime is } \tau_S) \tag{20}$$

For the medium and long intervals we also need to take into account the "starting point" of these, and we define:

$$P_{i,L,m}(t) = \Pr(Y(t) = y_i \cap \text{ current regime is } \tau_L \cap \text{ cycle is } m) \tag{21}$$

With "cycle" m we mean that the inspection will take place at point of times $\tau_L + (m-1)\tau_S, 2\tau_L + (m-1)\tau_S, 3\tau_L + (m-1)\tau_S, \ldots$. We can imagine that these cycles are running in parallel. In reality it will be only one cycle that could be active, but which one is actually active depends on when the system is renewed. Similarly we define:

$$P_{i,M,n}(t) = \Pr(Y(t) = y_i \cap \text{ current regime is } \tau_M \cap \text{ cycle is } n) \tag{22}$$

In total we have one $\mathbf{P}_S(t)$-vector, k_M $\mathbf{P}_{M,n}(t)$-vectors and k_L $\mathbf{P}_{L,m}(t)$-vectors. As we integrate the Markov differential equations for $t \neq l\tau_S$ we update all the $\mathbf{P}(t)$-vectors according to Equation (5).

Here it should be noted that a more efficient approach would be to use:

$$\mathbf{P}(t + \tau_S) \approx \mathbf{P}(t)[\mathbf{A}\Delta t + \mathbf{I}]^{2^n} \tag{23}$$

where $\Delta t = \tau_S/2^n$ and n is sufficient large to get a low value of Δt. Typically $n = 10$. This means that we may calculate $[\mathbf{A}\Delta t + \mathbf{I}]^{2^n}$ once, and use this matrix for all integrations. An alternative is to use matrix exponentials if we have the numerical routine available, i.e., $\mathbf{P}(t + \tau_S) = \mathbf{P}(t)e^{\mathbf{A}\tau_S}$.

Exercise 6 - Numerical precision

Assume $r = 5$ and $\lambda_i = 0.01, i = 0, 1, \ldots$ (time unit weeks). Assume the system is inspected with intervals of length $\tau = 26$, i.e., not varying intervals. Assume that the system is in state 0 at time $t = 0$. Find $\mathbf{P}(\tau)$ by using Equation (23) by using $n = 4, 6, 8$ and 10. What would be a reasonable value of n. □

Initially we have $P_{0,L,1}(t = 0)$, and all other probabilities are equal to zero. We now apply Equation (23) for $t = 0, \tau_S, 2\tau_S, 3\tau_S, \ldots$ for all the $\mathbf{P}(t)$-vectors. At each step we investigate each $\mathbf{P}(t)$-vector with respect to:

- Count the "number" of failures, to update the effective failure rate λ_E

- Count the "number" of repairs, i.e., update the renewal rate ρ_E
- Move "probability mass" to reflect repairs and change of inspection regime

For each $\mathbf{P}(t)$-vector we investigate the element corresponding to the failed state. These probabilities are added to a variable holding the accumulated expected number of failures. A failure could have occurred anywhere in the interval we integrate by Equation (23), and we assume an immediate repair. However, according to our assumptions, the system will not change the point of times where inspections are possible. Now consider time $t = l\tau_S$, then there will be a maintenance regime, say m with inspection interval τ_L which also has an inspection at time $t = l\tau_S$. Let p_j be all the probabilities representing failures in the set of $\mathbf{P}(t)$-vectors. The probabilities are now moved according to:

$$P_{0,L,m}(t^+) = P_{0,L,m}(t^-) + \sum_j p_j \tag{24}$$

where we assume that τ_L is small. Since the failure could have taken place some time before t, it is a probability that the system was reset to an as good as new state, and then jumped to the next deterioration level if τ_L is large. If this is the case we could split $\sum_j p_j$ to state 0 and 1.

We now proceed to handle the *change* of maintenance regime. As before we identify the integer value m which is such that the regime τ_L with cycle m has due date for an inspection at time $t = l\tau_S$. Similarly we identify the integer value n which is such that the regime τ_M with cycle n has due date for an inspection at time $t = l\tau_S$.

To understand the situation, consider $\tau_L = 6, \tau_M = 3$ and $\tau_L = 1$. Assume we are considering an inspection at time $t = l\tau_S = 13 \cdot 1 = 13$. m will now be 2 since the second τ_L cycle will have an inspection at times $1, 7, \mathbf{13}, 19 \ldots$. Further $n = 2$ because the second τ_M cycle will have an inspection at times $1, 4, 7, 10, \mathbf{13}, 16 \ldots$.

First consider $\mathbf{P}_{L,m}(t)$, i.e., the vector representing cycle m for the regime τ_L. For all states $i \in \mathscr{L}_L$ there will be no change in the inspection regime. For all states $i \in \mathscr{L}_M$ this will correspond to shifting from regime τ_L to regime τ_M. The vector $\mathbf{P}_{M,n}(t)$ is now representing the cycle which will "take over". Further, for all states $i \in \mathscr{L}_S$ this will correspond to shifting from regime τ_L to regime τ_S. Finally for all states $i \in \mathscr{L}_R$ this will correspond to a repair. We thus have:

$$\begin{aligned} P_{i,M,n}(t^+) &= P_{i,M,n}(t^-) + P_{i,L,m}(t^-), i \in \mathscr{L}_M \\ P_{i,S}(t^+) &= P_{i,S}(t^-) + P_{i,L,m}(t^-), i \in \mathscr{L}_S \\ P_{0,L,m}(t^+) &= P_{0,L,m}(t^-) + \sum_{i \in \mathscr{L}_R} P_{i,L,m}(t^-) \\ P_{i,L,m}(t^+) &= 0, i \notin \mathscr{L}_L \end{aligned} \tag{25}$$

The notation t^- and t^+ is used to denote time just before and just after an inspection respectively.

Referring to the example this means that if there is an inspection at time $t = 13$ there is a τ_L regime with cycle $m = 2$ which has an inspection at time $t = 13$. If it during the inspection is observed that the system is in $\mathscr{L}_M$, i.e., we find positive probabilities for $P_{i,L,m=2}(t^-), i \in \mathscr{L}_M$ we shift to a τ_L regime with cycle $n = 2$, i.e., inspection at time $t = 13$ and next inspection time at $t = 13 + 3 = 16$.

Next consider $\mathbf{P}_{M,n}(t)$. For all states $i \in \mathscr{L}_M$ there will be no change in the inspection regime. For all states $i \in \mathscr{L}_S$ this will correspond to shifting from regime τ_M to regime τ_S. The vector $\mathbf{P}_S(t)$ is now representing the regime which will "take over", i.e.,

$$\begin{aligned} P_{i,S}(t^+) &= P_{i,S}(t^-) + P_{i,M,n}(t^-), i \in \mathscr{L}_S \\ P_{0,L,m}(t^+) &= P_{0,L,m}(t^-) + \sum_{i \in \mathscr{L}_R} P_{i,M,n}(t^-) \\ P_{i,M,m}(t^+) &= 0, i \notin \mathscr{L}_M \end{aligned} \tag{26}$$

Finally consider $\mathbf{P}_S(t)$. For all states $i \in \mathscr{L}_S$ there will be no change in the inspection regime. For all states other states this will correspond to a repair to an good as new state. Note that for this regime there is not possible to be in $\mathscr{L}_L$ or $\mathscr{L}_M$. The updating of probabilities is defined by:

$$\begin{aligned} P_{0,L,m}(t^+) &= P_{0,L,m}(t^-) + \sum_{i \in \mathscr{L}_R} P_{i,S}(t^-) \\ P_{i,S}(t^+) &= 0, i \in \mathscr{L}_R \end{aligned} \tag{27}$$

Varying intervals for inspection - Alternative approach

In the previous section the Markov differential equations were integrated having all possible combination of sequences in the various $\mathbf{P}(t)$ vectors. Alternatively we could integrate the differential equations but when there is a failure or a demand for a renewal we just remove the corresponding probability mass from the $\mathbf{P}(t)$ vectors. This will reduce the number of $\mathbf{P}(t)$ vectors to consider.

We still assume that the time intervals between inspections are either τ_L, τ_M, or τ_S. Further k_L and k_M are integers such that $\tau_L = k_L \tau_S$ and $\tau_M = k_M \tau_S$. The situation now simplifies because there is only one τ_L regime and one τ_M regime, i.e., we are not considering the cycles any more. The time dependent probability vectors for each regime is given by $\mathbf{P}_L(t)$, $\mathbf{P}_M(t)$ and $\mathbf{P}_S(t)$.

In the integration we now define $\mathbf{h}$ to be a vector of probabilities of a failure in each period of length τ_S. Similarly $\mathbf{g}$ is a vector of probabilities of a request of a renewal in each period. Let j be an index running through all intervals of length τ_S. The $\mathbf{P}$-vector elements are defined by $P_{0,L}(t = 0) = 1$ and 0 for all other elements in the set of $\mathbf{P}(t)$-vectors. The integration

procedure is now:
For $j = 1, 2, \dots$ integrate all $\mathbf{P}(t)$-vectors according to:

$$\mathbf{P}(t + \tau_{\mathrm{S}}) \approx \mathbf{P}(t)[\mathbf{A}\Delta t + \mathbf{I}]^{2^n} \tag{28}$$

where $\Delta t = \tau_{\mathrm{S}}/2^n$, and where we update $t = t + \tau_{\mathrm{S}}$. In the formulas that follow t^- represents the time just prior to t, and t^+ represents the time just after t, i.e., when adjusting for the decision to take at time t.

Collect failure probabilities etc.:

$$h(j) = \sum_{k \in \{\mathrm{L,M,S}\}} P_{r,k}(t^-) \tag{29}$$

$$g(j) = \sum_{i \in \mathscr{L}_{\mathrm{R}}} P_{i,\mathrm{S}}(t^-), \text{ if } j \geq k_L \tag{30}$$

If $(j-1) \mod k_M = 0$ and $j > k_L$ then ("medium" maintenance):

$$g(j) = g(j) + \sum_{i \in \mathscr{L}_{\mathrm{R}}} P_{i,\mathrm{M}}(t^-) \tag{31}$$

$$P_{i,\mathrm{S}}(t^+) = P_{i,\mathrm{S}}(t^-) + P_{i,\mathrm{M}}(t^-), i \in \mathscr{L}_{\mathrm{S}} \tag{32}$$

End If
If $(j-1) \mod k_L = 0$ then ("long term" maintenance):

$$g(j) = g(j) + \sum_{i \in \mathscr{L}_{\mathrm{R}}} P_{i,\mathrm{L}}(t^-) \tag{33}$$

$$\begin{aligned} P_{i,\mathrm{S}}(t^+) &= P_{i,\mathrm{S}}(t^-) + P_{i,\mathrm{L}}(t^-), i \in \mathscr{L}_{\mathrm{S}} \\ P_{i,\mathrm{M}}(t^+) &= P_{i,\mathrm{M}}(t^-) + P_{i,\mathrm{L}}(t^-), i \in \mathscr{L}_{\mathrm{M}} \end{aligned} \tag{34}$$

End If
In the procedure listed above we have not explicitly "removed" the probability mass corresponding to a "request" for renewal. This we have to do.

The vectors **h** and **g** represent the expected number of failures and the number of requested renewals for each interval. Let **w** be the vector of the total expected number of renewals in each interval. For interval number 1 we have $w(1) = g(1)$, for interval number 2 the expected number of renewals is $w(2) = h(2) + w(1)g(1)$. To find the number of expected renewals in general we could use the discrete version of the renewal density for period j:

$$w(j) = g(j) + \sum_{i=1}^{j-1} w(j-i)g(i) \tag{35}$$

Since we already have calculated the values in **g** it is straight forward to obtain **w** from Equation (35).

Note that $w(j-i)g(i)$ represents the probability that there was a renewal at $(j-i)\tau_S$ and then there is another renewal $i\tau_S$ later. Here we should also account for the possibility that it was a *failure* at $(j-i)\tau_S$ and then there is another renewal $i\tau_S$ later, hence we more correct would be:

$$w(j) = g(j) + \sum_{i=1}^{j-1} [w(j-i) + f(j-i)]g(i) \tag{36}$$

where $f()$ is described below.

Let $\mathbf{f}$ be the vector of expected number of failures in each period. A failure will occur in period j in two disjoint ways, either the initial system fails in interval j, or there was a renewal or failure in a previous period $j-i$, and then this system fails after another i periods:

$$f(j) = h(j) + \sum_{i=1}^{j-1} [w(j-i) + f(j-i)]h(i) \tag{37}$$

Equation (37) may now be used to find the average effective failure rate over a given time horizon.

Exercise 7

Implement the model above, where $r = 5$, $\lambda_i = 0.01$, the maintenance rule is $\tau_L = 52, \tau_M = 26, \tau_S = 18, \mathscr{L}_L = \{0\}, \mathscr{L}_M = \{1,2\}, \mathscr{L}_S = \{3\}$ and $\mathscr{L}_R = \{4\}$. □

Varying intervals for inspection when inspection causes damage

In some situations an inspection may cause some damage to the system. Vatn and Pedersen (2020) present a conceptual model for degradation caused by inspections where we still may argue that the Markovian properties holds between inspections. In their model the failure rate after the n'th inspection is give by:

$$\lambda_n = \frac{\rho}{1 + m_0 e^{-\alpha n}} \tag{38}$$

where ρ is a rate of "external shocks", m_0 is an expression of the initial resistance against such shocks, and α is a measure of relative internal damage caused by each inspection.

There is no obvious way to extend this model to a multi-state model. Two extreme approaches could be:

1. The accumulated number of inspections, n, influence the transition rates independent on which states have been visited, i.e., $\lambda_{i,n} = \frac{\rho_i}{1+m_{0,i}e^{-\alpha_i n}}$
2. It is the number of inspections, n_i, conducted while being in state i which determines the transition rate, i.e., $\lambda_{i,n} = \frac{\rho_i}{1+m_{0,i}e^{-\alpha_i n_i}}$

In the following we will take the first approach as basis. Further the approach simplifies if ρ, m_0 and α are the same for all states i. This means that the transition matrix, $\mathbf{A}$ will depend on the number of inspections. After n inspection the transition matrix is then given by $\mathbf{A}_n$, and should be rather easy to establish if we are able to specify ρ, m_0 and α. If n is known, it is straight forward to update the time dependent solution:

$$\mathbf{P}(t+\tau_\mathrm{S}) \approx \mathbf{P}(t)[\mathbf{A}_n \Delta t + \mathbf{I}]^{2^m} = \mathbf{P}(t)\mathbf{E}_n \tag{39}$$

where $\Delta t = \tau_\mathrm{S}/2^m$ and m is sufficient large. Here $\mathbf{E}_n$ is an integration matrix approximating $e^{\mathbf{A}_n \tau_\mathrm{S}}$. When integrating the Markov equations according to the procedure described in the previous section we used Equation (39) for all $\mathbf{P}$-vectors with the same integration matrix $\mathbf{E}$. Now this is more complicated. It is convenient to calculate $\mathbf{E}_n, n = 0, 1, \ldots$ in advance before the integration process starts since these matrices are repeatedly needed. Then we need to keep track of n for each $\mathbf{P}$-vector. For $\mathbf{P}_\mathrm{L}$ this is straight forward since $n = \mathrm{int}\, j/k_L$. For $\mathbf{P}_\mathrm{M}$ it is more challenging. For inspections following the medium regime the number of inspections is a stochastic variable.

Introduce a vector $\mathbf{q}$ where element $q(j)$ is the probability that the medium regime was introduced in period j. As we integrate $q(j)$ is given by $q(j) = \sum_{i \in \mathscr{L}_\mathrm{M}} P_{i,\mathrm{L}}(t^-)$ following the notation in Equation (34). For a given j we may process the $\mathbf{q}$-vector up to element j to find the probability that it has been exactly $0, 1, 2, \ldots$ inspections up to the current period. Let these probabilities be denoted $\pi_0, \pi_1, \pi_2, \ldots$ respectively. The update of the time dependent solution is now:

$$\mathbf{P}_\mathrm{M}(t+\tau_\mathrm{S}) = \sum_{i=0,1,2,\ldots} \pi_i \mathbf{P}_\mathrm{M}(t)\mathbf{E}_i \tag{40}$$

For $\mathbf{P}_\mathrm{S}$ it is even more challenging to keep track of the number of inspections. A shift to the "small" interval regime could either be a result of an inspection from the medium regime or from the long regime. We therefore need to keep track of both of these two shifts. Let $\mathbf{r}$ be a vector where element $r(j)$ is the probability that the short interval regime was introduced in period j directly from the long interval regime, and $\mathbf{s}$ be a vector where element $s(j)$ is the probability that the short interval regime was introduced in period j from the medium interval regime.

For a given j we may process the $\mathbf{q}, \mathbf{r}$ and $\mathbf{s}$ vectors up to element j to find the probability that it has been exactly $0, 1, 2, \ldots$ inspections up to the current period given we are in the small interval regime. Let these probabilities be denoted $\pi_0, \pi_1, \pi_2, \ldots$ respectively. The update of the time dependent solution is now:

$$\mathbf{P}_\mathrm{S}(t+\tau_\mathrm{S}) = \sum_{i=0,1,2,\ldots} \pi_i \mathbf{P}_\mathrm{S}(t)\mathbf{E}_i \tag{41}$$

The entire procedure is now similar to the previous section. That is, Equations (29) to (37) are identical. The only difference is that the integration of the solutions in Equation (28) is replaced by the integration procedures described above.

Exercise 8

Work out the formulas for $\pi_0, \pi_1, \pi_2, \ldots$ in the two situations above. □

References

Vatn, J. and F. Pedersen (2020). An external shock - internal barrier degradation model to account for operational loads. In P. Baraldi, F. D. Maio, and E. Zio (Eds.), *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference*.