A Maintenance Model for Systems with Phase-type Distributed Times to Failure

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von	Dipl.-Wirtsch.-Math. Christoph Riethmüller
geb. am	08.09.1980 in Hamburg

genehmigt durch die Fakultät für Mathematik

der Otto-von-Guericke-Universität Magdeburg

Gutachter:	Prof. Dr. Gerd Christoph
	Prof. Dr. Nikolaos Limnios

eingereicht am:	11.01.2010
Verteidigung am:	29.03.2010

Contents

List of Figures

Summary

In this thesis we examine the failures of a system with phase-type distributed times to failure. For this purpose we consider a finite homogeneous Markov chain whose states are either transient or absorbing. We describe the stochastic failure rate of the system by the developing of the failure intensities of the Markov chain. This approach allows a straightforward interpretation: We assume that the system may work in a finite number of different conditions. These different conditions are the transient states of the Markov chain. While the system is working in one condition, the stochastic failure rate is constant. Changes of the system's condition are caused by random shocks, which are given by the jumps of the Markov chain. The time to failure of the system is the time to absorption of the Markov chain. The distribution of the time to absorption is a phase-type distribution.

We introduce different failure types which are represented by the absorbing states of the Markov chain. By this, minor and major failures may be modeled. In the main part of the thesis, we describe the user's opportunities of interaction. Having a failure, the system may be repaired. Furthermore, while the system is working, we may perform a preventive maintenance action. The maintenance actions are represented by stochastic matrices which, as well as the intensity matrix of the Markov chain, are the essential mathematical objects investigated. Within this approach, we cover the classical repair policies *minimal repair*, *replacement* and *imperfect repair*. Additionally, we introduce a preventive maintenance policy which suits the model and seems to be more feasible than the classical policies *block replacement* and *age replacement*. We also try to give a satisfying explanation to the question how to choose an optimal maintenance policy. The optimality criterion considered is to maximize the expected gains rewarded by the system. We discuss four approaches for solving the maximization problem, and we provide source codes for the implementation of these approaches in the statistical software R. We finish with a discussion of possible extensions of the model and subjects that may interest further research about this model.

Zusammenfassung

In dieser Arbeit betrachten wir das Ausfallverhalten eines Systems, dessen Ausfallzeitpunkt phase-type-verteilt ist. Dazu betrachten wir eine endliche homogene Markov-Kette, deren Zustände entweder transient oder absorbierend sind. Die stochastische Ausfallrate des Systems wird durch den Verlauf der Ausfallintensitäten der Markov-Kette beschrieben. Wir können diesen Ansatz wie folgt interpretieren: Wir nehmen an, dass das System in endlich vielen unterschiedlichen Zuständen arbeiten kann. Diese Zustände sind die transienten Zustände der Markov-Kette. Während das System in einem dieser Zustände ist, bleibt die stochastische Ausfallrate konstant. Änderungen des Zustandes und damit auch Änderungen der stochastischen Ausfallrate erfolgen durch zufällige Schocks. Diese Änderungen sind die Sprünge der Markov-Kette. Der Ausfallzeitpunkt des Systems ist der Absorptionszeitpunkt der Markov-Kette. Die Verteilung der Absorptionszeit ist eine phase-type-Verteilung.

Die absorbierenden Zustände der Markov-Kette stellen unterschiedliche Ausfallarten dar. So können wir geringfügige und schwerwiegende Ausfälle beschreiben. Im Hauptteil der Arbeit betrachten wir die Handlungsmöglichkeiten eines Anwenders des Systems. Nach einem Ausfall kann das System durch Reparatur wieder in einen arbeitsfähigen Zustand versetzt werden, und, während das System arbeitet, können vorbeugende Instandhaltungsmaßnahmen durchgeführt werden. Die Instandhaltungsmaßnahmen werden durch stochastische Matrizen dargestellt. Wir untersuchen das Zusammenspiel dieser stochastischen Matrizen mit der Intensitätsmatrix der Markov-Kette. Diese Herangehensweise beinhaltet die klassischen Reparaturstrategien *minimale Reparatur*, *Erneuerung* und *unvollständige Reparatur*. Zudem führen wir eine vorbeugende Instandhaltungsstrategie ein, welche die Eigenschaften unseres Modells ausnutzt und anscheinend brauchbarer ist als die klassischen Strategien *Blockerneuerung* und *altersbedingte Erneuerung*. Wir versuchen außerdem befriedigend zu erklären, wie man eine optimale Instandhaltungsstrategie auswählen kann. Als Optimalitätskriterium betrachten wir die erwarteten Gewinne, die vom System erwirtschaftet werden. Wir betrachten vier Verfahren zur Lösung dieses Problems und stellen Quellcodes zur Implementierung dieser Verfahren in dem Statistiksoftwareprogramm R zur Verfügung. Die Arbeit endet mit einer Diskussion über mögliche Erweiterungen des Modells, sowie Ansätze für weitere Forschungen.

Used symbols and acronyms:

$\mathbb{N}$	the set of integers, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$
$\mathbb{R}$	the set of real numbers, $\mathbb{R}^+ = (0,\infty)$, $\mathbb{R}_0^+ = [0,\infty)$, $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, $\overline{\mathbb{R}^+} = \mathbb{R}^+ \cup \{+\infty\}$, $\overline{\mathbb{R}_0^+} = \mathbb{R}_0^+ \cup \{+\infty\}$
$\mathbb{C}$	the set of complex numbers
$\forall, \exists, \exists!$	$\forall$ for all, $\exists$ exist/-s, $\exists!$ exists exactly one
$x \in A$	x is an element of the set A, $\subset$ subset, $\subseteq$ subset or equal
$A \cup B$	union, $A \cap B$ intersection of sets A, B, $\overline{A}$ complement of A
$\mathfrak{P}(A)$	power set (set of all subsets) of a set A
δ_{ij}	Kronecker-symbol: $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$
0	null-matrix $0 \in \mathbb{R}^{n\times m}$, $n, m \in \mathbb{N}$ (includes null-vector and number 0)
e_i	unit vectors $v \in \mathbb{R}^n$: $v_j = \delta_{ij}$, $j = 1, \ldots, n$
$\mathbf{1}$	one-vector $v \in \mathbb{R}^n$: $v_i = 1$, $i = 1, \ldots, n$
$\mathbf{I}$	identity matrix $M \in \mathbb{R}^{n\times n}$: $M_{ij} = \delta_{ij}$, $i, j = 1, \ldots n$
$\mathrm{diag}(a)$	diagonal matrix with diagonal entries $a \in \mathbb{R}^n$: $(\mathrm{diag}(a))_{ij} = a_i \delta_{ij}$
$\mathrm{tr}(A)$	trace of a quadratic matrix A
$\det(A)$	determinant of a quadratic matrix A
$\lfloor x \rfloor$	integer part of a nonnegative real number x[1]
$\mathbb{1}_A(x)$	indicator function: $\mathbb{1}_A(x) = \begin{cases} 1 & \text{, if } x \in A \\ 0 & \text{, else} \end{cases}$
$o(f(x))$	Landau-symbol[2]
$x \to x_0+$	right-hand convergence: $x \to x_0, x > x_0$, also $f(x_0+)$
$x \to x_0-$	left-hand convergence: $x \to x_0, x < x_0$, also $f(x_0-)$
iid	independent, identically distributed
iff	if and only if
a.s.	almost surely[3]
wlg	without loss of generality
cadlag	continu à droit limité à gauche (right-continuous and left-hand limits)
SSM	smooth semimartingale
MC	Markov chain
cdf	cumulative distribution function

[1] $\lfloor x \rfloor = \sup\{n \in \mathbb{N}_0 : n \leq x\}$, $x \geq 0$.

[2] We write $a(x) = o(f(x))$ for $x \to x_0 \in \overline{\mathbb{R}}$ if $\lim\limits_{x\to x_0} \dfrac{a(x)}{f(x)} = 0$

[3] An assertion $\mathcal{A}$ holds P-a.s. iff $\mathcal{A}$ is true on a set A with $\mathrm{P}(A) = 1$

Chapter 1

Introduction

In this thesis we examine the failures of a system, e.g. breakdowns of an engine or diseases of a human being. We assume that the time to failure of the system follows a probability distribution on the non-negative real line. Before failure the proneness to failure is described by the failure rate $\lambda(t)$ of this distribution. Choosing an appropriate failure rate is the main aspect in modeling.

We consider phase-type distributed times to failure. The class of phase-type distributions was introduced by Neuts [45] as a generalization of the Erlang method of stages. A phase-type distribution is the distribution of the time to absorption τ of an absorbing finite homogeneous time-continuous Markov chain J. Roughly spoken, the phase-type distributions are all mixtures of convolutions of exponential distributions. They have a couple of appealing properties (cf. Neuts [47] or O'Cinneide [51, 52]), and all the same they allow a good mathematical tractability. There are many approaches to apply them in areas concerning waiting times, killing times and times to failure, e.g. in queuing theory (Neuts [46]), risk theory (Asmussen [6], Bladt [16], Jacobsen [30]), survival analysis (Aalen [2]) and reliability theory (Neuts and Meier [48]).

The failure rate of the system may now be represented by the absorption rate of the underlying Markov chain. In fact, the current absorption rate of the Markov chain $\overline{\lambda}(t)$ is the conditional failure rate of the system given the current state of the Markov chain. While the (unconditional) failure rate $\lambda(t)$ is a rational function, the conditional failure rate $\overline{\lambda}(t)$ is a piecewise constant jump process, hence we have an intuitional interpretation of the latter. A random variable with a failure rate which depends on an underlying random variable or stochastic process, like τ in our model, is called doubly stochastic. For more on doubly stochastic random variables see Cox [21] (regression model), Aalen [1] (multiplicative hazard model) or Finkelstein

[26] (a general review).
Maintenance actions are the opportunities to interact provided by the system. We distinguish between repair actions, which are performed in case of a failure of the system, and preventive maintenance actions, which are performed as long as the system is working. For the common repair policies see Beichelt [10] (minimal repair), Brown and Proschan [20] (imperfect repair), and for the common preventive maintenance policies see Barlow and Proschan [9], Bosch [19], Jensen [32], Beichelt [11] or Aven and Jensen [8].

The aim of this thesis is to introduce a maintenance model for systems with phase-type distributed times to failure. We want to introduce maintenance policies which are suitable for the model of the times to failure, and we want to give satisfying approaches for the optimization of the maintenance policy.
The structure of the thesis is as follows.
In Chapter 2 we present the mathematical framework of the thesis. We define basic terms of stochastic calculus and survival analysis. We derive a method for the calculation of matrix exponentials by using matrix calculus from linear algebra. We give a deeper introduction into homogeneous Markov chains, where the matrix exponentials appear as transition matrices. Afterwards, phase-type distributions are defined and motivated. Next, we introduce martingales and compensator, especially applied for counting processes. We present an algorithm for generating a distribution on a finite support. In this chapter we jump between different topics. But once we read this chapter, we are prepared for the maintenance model presented later.
In Chapter 3 we define a failure model. We introduce the working states and the failure states of the system. We investigate the obtained failure rate and show that the conditional failure rate equals the canonical intensity process of the failure indicator process $(\mathbb{1}_{[0,t]}(\tau))_{t\geq 0}$.
Chapter 4 deals with maintenance actions in case of a failure. We represent a repair action by a stochastic matrix R. We show that, if we relaunch the system after a failure instantaneously with respect to the repair matrix R, the state of the system is again driven by a homogeneous Markov chain S^R. We derive the expected number of failures in dependence of R, and we use them to choose an optimal repair matrix. The optimality criterion considered is to maximize the expected gains. We discuss four approaches in order to optimize the expected gains. First, we calculate the expected gains exactly, here we need the matrix exponentials from Chapter 2. Second, we calculate the limiting behavior of S^R in order to obtain a repair matrix R which is finally optimal. The third approach is empirical. We simulate the gains for the different repair policies, here we use the alias algorithm

presented in Chapter 2. The last approach is a heuristic one. Similar to dynamic programming we choose the repair matrix R which yields the best results in a K-step algorithm.
In Chapter 5 we introduce a preventive maintenance policy P. Within this policy a preventive maintenance action is performed, if the system enters a critical working state. We discuss the behavior of the state of the system in dependence of P. Again, we may use the four approaches from Chapter 4 for choosing an optimal maintenance policy (R, P).
In Chapter 6 we give some concluding remarks, and we highlight possible topics for further research. Especially, we show up possibilities for statistical inference.
We clarify the model and the methods introduced with the help of an example, which is presented throughout the Chapters 3 to 5.
We will provide additional informations for the interested reader by using footnotes.
We conclude the introduction with some notations. Throughout the thesis a vector v is always a column vector, the respective row vector is v'. The dimensions of vectors and matrices are omitted if they are obvious. We use the further conventions $\inf \emptyset := \infty$ and $\sum_{\emptyset} \cdots := 0$. The upright P denotes the probability measure and the italic P denotes a preventive maintenance matrix.

Chapter 2

Fundamentals and tools

2.1 Basic definitions of stochastic calculus

In this section we introduce basic terms and definitions of stochastic calculus which should be familiar to undergraduates having attended first courses in probability and stochastic processes. The main purpose is having a clear nomenclature throughout the thesis.

Definition 2.1. Let Ω be a non-empty set.

i) A *σ-algebra* $\mathfrak{A}$ over Ω is a collection of subsets of Ω satisfying

   - $\Omega \in \mathfrak{A}$,
   - $\mathfrak{A}$ is closed under complements:

$$A \in \mathfrak{A} \Rightarrow \overline{A} \in \mathfrak{A}, \ \forall \ A \in \mathfrak{A},$$

   - $\mathfrak{A}$ is closed under countable unions:

$$A_1, A_2, ... \in \mathfrak{A} \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathfrak{A}.$$

   The elements of a σ-algebra are called events and the pair $(\Omega, \mathfrak{A})$ is called *measure space*.

ii) A *probability measure* P on a σ-algebra $\mathfrak{A}$ is a function $\mathrm{P} : \mathfrak{A} \to [0, 1]$ satisfying

   - $\mathrm{P}(\Omega) = 1$,
   - $A_1, A_2, ... \in \mathfrak{A}$ with $A_i \cap A_j = \emptyset$ if $i \neq j$

$$\Rightarrow \mathrm{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mathrm{P}(A_i) \ .$$

The triple $(\Omega, \mathfrak{A}, \mathrm{P})$ is called *probability space*.

iii) For two events $A, B \in \mathfrak{A}$ with $\mathrm{P}(B) > 0$ we define the conditional probability of the event A given event B via

$$\mathrm{P}(A|B) := \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \ .$$

The function $\mathrm{P}_B : \mathfrak{A} \to [0,1]$ with $\mathrm{P}_B(A) := \mathrm{P}(A|B)$ is a probability measure on $\mathfrak{A}$.

iv) The events $(A_i)_{i \in I}$ (I is an arbitrary index set) are called *independent* if for all finite subsets B of I it holds

$$\mathrm{P}\left(\bigcap_{i \in B} A_i\right) = \prod_{i \in B} \mathrm{P}(A_i) \ .$$

The events are called *pairwise independent* if for any two distinct elements $i, j \in I$ it holds

$$\mathrm{P}(A_i \cap A_j) = \mathrm{P}(A_i)\mathrm{P}(A_j) \ .$$

Every σ-algebra $\mathfrak{A}$ over Ω satisfies

$$\{\emptyset, \Omega\} \subseteq \mathfrak{A} \subseteq \mathfrak{P}(\Omega)$$

and $\{\emptyset, \Omega\}$, $\mathfrak{P}(\Omega)$ are σ-algebras. The standard choice for a σ-algebra over a topological space[1] Ω is the Borel[2]-σ-algebra $\mathfrak{B}(\Omega)$ which is generated by the topology (the open subsets) of Ω.

Definition 2.2. Let $(\Omega, \mathfrak{A}, \mathrm{P})$ be a probability space, and let $(E, \mathfrak{E})$ be a measure space. A *random variable* X is a mapping $X : \Omega \to E$ that is $\mathfrak{A}$-$\mathfrak{E}$-measurable, i.e.

$$X^{-1}(B) \in \mathfrak{A} \ \forall \ B \in \mathfrak{E} \ .$$

We henceforth assume a universal probability space $(\Omega, \mathfrak{A}, \mathrm{P})$ for all random variables considered and we assume that $\mathfrak{A}$ is complete (if $A \in \mathfrak{A}$ with $\mathrm{P}(A) = 0$ and $B \subset A$, then also $B \in \mathfrak{A}$). The respective measure space $(E, \mathfrak{E})$ of a random variable X will be referred to by calling X $(E, \mathfrak{E})$-valued. If we consider the Borel-σ-algebra $\mathfrak{E} = \mathfrak{B}(E)$ we just note that X is E-valued. An $\mathfrak{A}$-$\mathfrak{B}(E)$-measurable mapping is called Borel-measurable.

[1] A topological space is a set Ω together with O, a collection of subsets of Ω, satisfying the following axioms:
- The empty set and Ω are in O.
- The union of any collection of sets in O is also in O.
- The intersection of any finite collection of sets in O is also in O.

[2] Named after Émile Borel (1871-1956), a French mathematician and politician.

Example 2.1. The indicator function $\mathbb{1}_A$ is Borel-measurable iff $A \in \mathfrak{A}$.

Remark 2.1. A random variable X induces a probability measure P^X on $(E, \mathfrak{E})$ by

$$\mathrm{P}^X(B) := \mathrm{P}(X^{-1}(B)) = \mathrm{P}(\{\omega \in \Omega : X(\omega) \in B\}) \ , \quad B \in \mathfrak{E} \ .$$

Instead of $\mathrm{P}^X(B)$ we will usually write $\mathrm{P}(X \in B)$. We call P^X the *distribution* of X. The set

$$\sigma(X) := \{X^{-1}(B) : B \in \mathfrak{E}\}$$

is a σ-algebra. It is called the σ-algebra induced by X and it is the smallest σ-algebra for which X is measurable[3].

The next lemma is Lemma 1.13 from Kallenberg [33]

Lemma 2.1. For $i = 1, 2$ let X_i be an E_i-valued random variable. Then X_1 is $\sigma(X_2)$-measurable iff there exists some measurable mapping $h : E_2 \to E_1$ with $X_1 = h(X_2)$.

Definition 2.3. For $\mathbb{R}$-valued random variables X, Y and functions $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R}^2 \to \mathbb{R}$ we define

i) the *cumulative distribution function*[4] (cdf)

$$F_X(x) := \mathrm{P}(X \leq x) \ , \ x \in \mathbb{R} \ .$$

ii)

$$\mathrm{E}(g(X)) := \int g(X) \mathrm{dP} \ ,$$

if the integral[5] exists. We call $\mathrm{E}(X)$ *mean*, $\mathrm{Var}(X) := \mathrm{E}((X - \mathrm{E}(X))^2)$ *variance* and $\mathrm{E}(X^k)$ k-th *moment* of X ($k \in \mathbb{N}$).

iii)

$$\mathrm{E}(h(X, Y)) := \int h(X, Y) \mathrm{dP} \ ,$$

if the integral exists. We call $\mathrm{Cov}(X, Y) := \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y)))$ *covariance* of X and Y.

[3] If X is $\mathfrak{A}_0$-$\mathfrak{E}$ -measurable for a σ-algebra $\mathfrak{A}_0$, then $\sigma(X) \subseteq \mathfrak{A}_0$

[4] The cdf F always satisfies
- F is right-continuous.
- F is monotone increasing.
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

[5] Integrals with respect to a measure P are matter of measure and integral theory and will not be discussed here, but see e.g. Billingsley [14] or Kallenberg [33]

Definition 2.4. Let X be a random variable with the cumulative distribution function F.

i) If there exists a non-negative function $f : \mathbb{R} \to [0, \infty)$ with

$$F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x \ , \ \forall \ t \in \mathbb{R} \ ,$$

we call (the distribution of) X *absolute continuous* and the function f is called *density function*[6] of X.

ii) If there exists a discrete set $A = \{a_1, a_2, ...\} \subset \mathbb{R}$ with $\mathrm{P}(X \in A) = 1$ and $\mathrm{P}(X = a_i) > 0 \ \forall \ i$, we call (the distribution of) X *discrete*. The set A is called *support* of X and the function $\mathrm{P}(X = x) \ , x \in A$ is called *probability function* of X.

Example 2.2. A random variable X is called *exponentially*-(λ)-distributed (short $X \sim \mathrm{Exp}(\lambda)$) if X is absolute continuous with the density function

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x) \ , \ x \in \mathbb{R} \ ,$$

for some $\lambda > 0$. The mean of the distribution is $\mathrm{E}(X) = 1/\lambda$.

Example 2.3. A random variable X is called *geometrically*-(p)-distributed (short $X \sim \mathrm{Geo}(p)$) if X is discrete with the probability function

$$\mathrm{P}(X = k) = p(1-p)^{k-1}, \ k \in \mathbb{N} \ ,$$

for some $p \in (0, 1]$. The mean of the distribution is $\mathrm{E}(X) = 1/p$.

These two distributions are connected in the following manner:

$$X \sim \mathrm{Exp}(\lambda) \ \Rightarrow \ \lfloor X \rfloor + 1 \sim \mathrm{Geo}(p = 1 - e^{-\lambda}) \ .$$

Example 2.4. A random variable X is called *uniformly*-distributed

i) on a finite set $\{x_1, ..., x_n\} \subset \mathbb{R}$ (short $X \sim \mathrm{U}(\{x_1, ..., x_n\})$), if

$$\mathrm{P}(X = x_i) = \frac{1}{n} \ , \ i = 1, ..., n \ .$$

ii) on an interval $[a, b] \subset \mathbb{R}$ with $a < b$ (short $X \sim \mathrm{U}([a, b])$), if X is absolute continuous with the density function

$$f(x) = (b-a)^{-1} \mathbb{1}_{[a,b]}(x) \ , \ x \in \mathbb{R} \ .$$

[6]The density function is not unique. If two functions f_1, f_2 satisfy Definition 2.4, then $f_1 = f_2$ P-a.s.

Definition 2.5. The random variables $(X_i)_{i\in I}$ (I is an arbitrary index set, X_i is $(E_i, \mathfrak{E}_i)$-valued) are called *independent* if for all finite subsets B of I holds

$$\mathrm{P}\left(\bigcap_{i\in B}\{X_i \in A_i\}\right) = \prod_{i\in B}\mathrm{P}(X_i \in A_i) \ ,$$

for all $A_i \in \mathfrak{E}_\mathfrak{i}$, $i \in B$. They are called *pairwise independent* if for any two distinct $i, j \in I$ holds

$$\mathrm{P}(X_i \in A_i, X_j \in A_j) = \mathrm{P}(X_i \in A_i)\mathrm{P}(X_j \in A_j) \ ,$$

for all $A_i \in \mathfrak{E}_i,\ A_j \in \mathfrak{E}_j$.

Remark 2.2. For $\mathbb{R}$-valued random variables $(X_i)_{i\in I}$ with respective cumulative distribution functions F_i independence is equivalent to

$$\forall\ n \in \mathbb{N}\ \forall\ i_1, ..., i_n \in I\ \forall\ x_1, ..., x_n \in \mathbb{R}:$$

$$F_{X_{i_1},...,X_{i_n}}(x_1, ..., x_n) := P(X_{i_1} \le x_1, ..., X_{i_n} \le x_n) = \prod_{k=1}^{n} F_{X_{i_k}}(x_k) \ .$$

The function $F_{X_{i_1},...,X_{i_n}}$ is called *joint distribution function* of $X_{i_1}, ..., X_{i_n}$.

Definition 2.6. Let T be a non-empty set. A family of $(E, \mathfrak{E})$-valued random variables $(X_t)_{t\in T}$ is called *stochastic process*. The set T is called index set and the set E is called state space.

We usually interpret a stochastic process as a random variable evolving in time starting in $t = 0$, so $T \subseteq [0, \infty)$. Alongside with this interpretation comes the definition of a filtration.

Definition 2.7. Let $\emptyset \neq T \subseteq [0, \infty)$. A family of σ-algebras $\mathfrak{F} = (\mathfrak{F}_t)_{t\in T}$ with $\mathfrak{F}_t \subseteq \mathfrak{A}\ \forall\ t \in T$ is called *filtration* if

$$\forall\ s < t:\ \mathfrak{F}_s \subseteq \mathfrak{F}_t\ ,\ s, t \in T \ .$$

We call the quadruple

$$(\Omega, \mathfrak{A}, (\mathfrak{F}_t)_{t\in T}, \mathrm{P})$$

a *filtered probability space*. A stochastic process $(X_t)_{t\in T}$ is called *adapted* to a filtration $\mathfrak{F}$ if

$$\forall\ t \in T:\ X_t \text{ is } \mathfrak{F}_t\text{-measurable} \ .$$

Let $\sigma(X_s, 0 \le s \le t)$ be the smallest σ-algebra for which all $X_s : 0 \le s \le t$ are measurable. For a stochastic process $(X_t)_{t\in T}$ the filtration $\mathfrak{F}^X = (\mathfrak{F}^X_t)_{t\in T}$ with

$$\mathfrak{F}^X_t := \sigma(X_s, 0 \le s \le t)\ ,\ t \in T$$

is called *canonical filtration* of $(X_t)_{t\in T}$ and, of course, $(X_t)_{t\in T}$ is adapted to $\mathfrak{F}^X$.

If a stochastic process $(X_t)_{t\in T}$ is adapted to a filtration $\mathfrak{F}$ we interpret $\mathfrak{F}_s$ to be the amount of knowledge about $(X_t)_{t\in T}$ up to the time s. We assume that all filtrations considered satisfy the usual conditions of right-continuity and completeness (cf. Aven and Jensen [8] Definition 33).

Definition 2.8. A $\overline{\mathbb{R}^+}$-valued stochastic process $(T_n)_{n\in\mathbb{N}}$ is called *point process*, if

i) for all $n \in \mathbb{N}$: $T_n \leq T_{n+1}$ P-a.s.

ii) for all $n \in \mathbb{N}$ with $\mathrm{P}(T_n < \infty) > 0 : \mathrm{P}(T_n < T_{n+1} | T_n < \infty) = 1$

iii) $\lim_{n\to\infty} T_n = \infty$ P-a.s.

T_n is called n-th *arrival time* and $\Delta_n := T_n - T_{n-1}$ (when defined and with $\Delta_1 := T_1$) is called n-th *waiting time*.

A point process is an unbounded monotone increasing series of random variables (strictly if finite). A typical application are failure times of a system.

Definition 2.9. Let $(T_n)_{n\in\mathbb{N}}$ be a point process and let $(U_n)_{n\in\mathbb{N}}$ be an E-valued stochastic process. The process $(T_n, U_n)_{n\in\mathbb{N}}$ is called *marked point process* and E is the set of marks.

Definition 2.10. For a point process $(T_n)_{n\in\mathbb{N}}$ we define the corresponding *counting process* $(N_t)_{t\geq 0}$ by

$$N_t := \sum_{n=1}^{\infty} \mathbb{1}_{[0,t]}(T_n) \quad , \ t \geq 0.$$

A counting process is a monotone increasing $\mathbb{N}_0$-valued stochastic process whose jump heights are equal to 1 P-a.s. For a counting process we obtain the corresponding point process: $T_n = \inf\{t \geq 0 : N_t = n\}$, $n \in \mathbb{N}$.

Definition 2.11. A point process is called *renewal process* if the waiting times $\Delta_1, \Delta_2, ...$ are iid. The corresponding counting process is called renewal counting process. If only the waiting times $\Delta_2, \Delta_3, ...$ are iid, the process is called delayed renewal process.

A common application of renewal processes are failure times of a system, if the system is repaired completely after failure (replacement). If we observe such a system starting from an arbitrary time, the first time to failure may have a different distribution than the following ones. Therefore the failure times may be modeled by a delayed renewal process.

Definition 2.12. A $\overline{\mathbb{R}_0^+}$-valued random variable τ on a filtered probability space $(\Omega, \mathfrak{A}, (\mathfrak{F}_t)_{t\in T}, \mathrm{P})$ is called *stopping time* if

$$\{\tau \leq t\} \in \mathfrak{F}_t \ \forall\ t \in T.$$

For a stochastic process $(X_t)_{t\geq 0}$ and a P-a.s. finite stopping time τ the random variable $X_\tau := X_{\tau(\omega)}(\omega)$ gives the value of the process at the stopping time.

Definition 2.13. Let X be a $\mathbb{R}$-valued random variable. The *characteristic function* $\varphi_X(t)$ of X is:

$$\varphi_X(t) := \int e^{itX} \mathrm{dP}\ , \quad t \in \mathbb{R}.$$

The characteristic function exists for all $t \in \mathbb{R}$ since $|\varphi_X(t)| \leq 1$.

Example 2.5. Let $X \sim \mathrm{Exp}(\lambda)$ with $\lambda > 0$. The characteristic function of X is

$$\varphi_X(t) = \int_0^\infty e^{itx}\lambda e^{-\lambda x}\mathrm{d}x = \lambda\left[(it-\lambda)^{-1}e^{(it-\lambda)x}\right]_0^\infty = \frac{\lambda}{\lambda - it}\ .$$

2.2 Classes of life-time distributions in survival analysis

In this section we introduce some non-parametric classes of distributions, that are often considered, when life-times[7] of a system are observed. An overview can be found in Beichelt [12] or Marshall and Olkin [40]. We focus only on two of the classes considered there, namely monotone failure rates and bathtub-shaped failure rates. First we introduce the failure rate.

Definition 2.14. Let $X \geq 0$ be an absolute continuous random variable with the distribution function F and the density function f. The function $\lambda : [0, \infty) \to [0, \infty]$ with

$$\lambda(x) = \begin{cases} \dfrac{f(x)}{1-F(x)} & ,\ x \geq 0,\ F(x) < 1\ , \\ 0 & ,\ x \geq 0,\ F(x) = 1\ . \end{cases}$$

is called failure rate.

[7] Life-times are usually non-negative, so we assume $X \geq 0$ P-a.s.

Since

$$f(x) = \lim_{\delta \to 0+} \frac{\mathrm{P}(x < X \leq x + \delta)}{\delta} \ , \ x \geq 0 \ ,$$

a straightforward interpretation of the failure rate is

$$\mathrm{P}(X \leq x + \delta | X > x) = \lambda(x)\delta + o(\delta) \ , \ x \geq 0 \ , \ F(x) < 1 \ . \tag{2.1}$$

The failure rate gives the proneness to failure of a system that is still working and is therefore an important function for modeling the failure behavior of a system. The distribution function may be calculated from the failure rate via the *exponential formula*

$$F(x) = 1 - \exp\left\{ - \int_0^x \lambda(u) \mathrm{d}u \right\} \ , \ x \geq 0 \ . \tag{2.2}$$

Changes of the failure rate may be interpreted in many applications as ageing or rejuvenating of the system. Ageing of a system may be caused by wear or stress and rejuvenation by maintenance actions or just by a change of the environment. We are especially interested in the following kinds of failure rates.

Definition 2.15. Let $\lambda(x)$ be the failure rate of an absolute continuous random variable X.

i) If $\lambda(x)$ is monotone increasing (decreasing) we say that X has an *increasing* failure rate (IFR) (resp. *decreasing* failure rate (DFR)) distribution.

ii) If $\lambda(x)$ is continuous and there exists $0 < x^* < \infty$ with

$$\lambda(x) \begin{cases} \text{monotone decreasing} & , x < x^* \\ \text{monotone increasing} & , x > x^* \end{cases} \ ,$$

we say that X has a *bathtub-shaped* failure rate.

The IFR (DFR) distributions are the most obvious classes of failure rate models. The bathtub models are often used if there's a high probability of an early failure because the system might start in a bad condition. The typical graph of a bathtub-shaped failure rate resembles the bathtub curve in Figure 2.1.[8]

[8]Figure 2.1 is from Wikipedia: en.wikipedia.org/wiki/Bathtub_curve

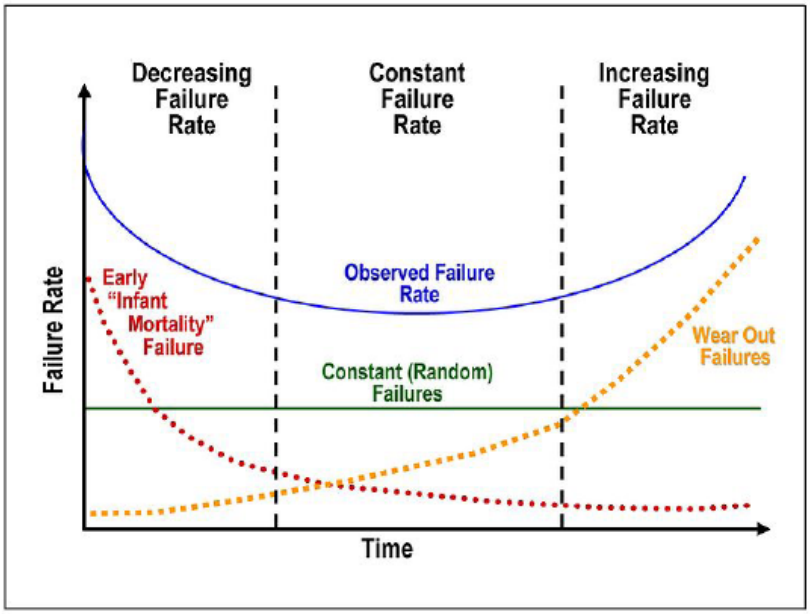


Figure 2.1: The bathtub curve

The definition of a bathtub-shaped failure rate is the one used by e.g. Block et al. [17].[9] For every parametric model it is an interesting question whether one may deduce the belonging to some failure rate class easily for given parameters.

2.3 Matrix exponentials

Let $A \in \mathbb{C}^{n\times n}$ be a square matrix. Then e^A is defined by the power series

$$e^A := \mathbf{I} + \sum_{k=1}^{\infty} \frac{A^k}{k!} \ .$$

Of course, the infinite sum converges for any matrix norm. There are several ways how to compute such a matrix-exponential, which have been subject to wide research. An overview of different methods was given by Moler and van Loan [43, 44]. Here we have a closer look on a method, which uses the

[9]There is a different common definition used by e.g. Marshall and Olkin [40] which assumes that the failure rate takes its minimum (and hence is constant) on an interval $[x^{*1}, x^{*2}]$. But this condition is not fulfilled by the usual parametric life-time models, including the one we are going to investigate.

eigenvalues of A. We implement this procedure in the statistic software R[10] and use the algorithm for the calculations in our examples. In this section we need basic terms from linear algebra. Let $A \in \mathbb{C}^{n\times n}$ be a square matrix. Similar to the notation by Agaev and Chebotarev [4] we denote by $\mathcal{R}(A)$ and $\mathcal{N}(A)$ the range respectively the null space of A and rank(A) gives the maximal number of linear independent row vectors of A. The polynomial

$$c_A(\lambda) = \det(\lambda \mathbf{I} - A) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{m_i} \quad , \; m_1 + ... + m_k = n \tag{2.3}$$

is the *characteristic polynomial* of A whose pairwise distinct zeros $\lambda_1, ..., \lambda_k$ are the eigenvalues of A with respective algebraic multiplicities $m_1, ..., m_k$.

Definition 2.16. Let $A \in \mathbb{C}^{n\times n}$ be a square matrix.

i) The *index* (or geometric multiplicity) ν of A is the smallest $k \in \mathbb{N}_0$ for which $\text{rank}(A^{k+1}) = \text{rank}(A^k)$ and we write $\nu = \text{ind}(A)$.

ii) Let λ be an eigenvalue of A and let $\nu_\lambda = \text{ind}(A - \lambda \mathbf{I})$. The *eigenprojection* Z_λ of A at λ is the unique idempotent matrix with

$$\mathcal{R}(Z_\lambda) = \mathcal{N}\left((A - \lambda \mathbf{I})^{\nu_\lambda}\right) \quad , \quad \mathcal{N}(Z_\lambda) = \mathcal{R}\left((A - \lambda \mathbf{I})^{\nu_\lambda}\right) \; .$$

The index ν of a matrix A equals zero iff A is regular. If the eigenvalue λ of A has the algebraic multiplicity m, then the index ν_λ is smaller or equal to m. We quote a classic result about operators on matrices, Theorem 8, Chapter VII in Dunford and Schwartz [25].

Theorem 2.1. Let A be a $n \times n$ matrix with characteristic polynomial

$$c_A(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{m_i} \quad , \; m_1 + ... + m_k = n \; ,$$

where $\lambda_1, ..., \lambda_k$ are the pairwise distinct eigenvalues of A with respective algebraic multiplicities $m_1, ..., m_k$ and indexes $\nu_i = \text{ind}(A - \lambda_i \mathbf{I})$, $i = 1, ..., k$. Let $f(x)$ be a complex function analytic in some open set containing $\{\lambda_1, ..., \lambda_k\}$ having the power series expansion $f(x) = \sum_{i=0}^{\infty} a_i x^i$. Then $f(A)$ can be defined as $f(A) = \sum_{i=0}^{\infty} a_i A^i$ and

$$f(A) = \sum_{i=1}^{k} \sum_{\alpha=0}^{\nu_i - 1} \frac{f^{(\alpha)}(\lambda_i)}{\alpha!} (A - \lambda_i \mathbf{I})^\alpha Z_{\lambda_i} \; ,$$

where Z_λ is the eigenprojection of A at λ.

[10] See http://www.r-project.org/

So if we know the matrices Z_λ we have

$$e^A = \sum_{i=1}^{k} \sum_{\alpha=0}^{\nu_i - 1} \frac{e^{\lambda_i}}{\alpha!} (A - \lambda_i \mathbf{I})^\alpha Z_{\lambda_i} \ . \tag{2.4}$$

Remark 2.3. Instead of a matrix A we might calculate the exponential of the matrix At with $t > 0$. If A has the eigenvalues $\lambda_1, ..., \lambda_k$ with respective indexes $\nu_1, ..., \nu_k$, then the matrix At has the eigenvalues $\lambda_1 t, ..., \lambda_k t$ with respective indexes $\nu_1, ..., \nu_k$.

We plug-in the matrix At in (2.4) and we obtain

$$e^{At} = \sum_{i=1}^{k} \sum_{\alpha=0}^{\nu_i - 1} \frac{e^{\lambda_i t}}{\alpha!} (At - \lambda_i t \mathbf{I})^\alpha Z_{\lambda_i t} = \sum_{i=1}^{k} \sum_{\alpha=0}^{\nu_i - 1} \frac{t^\alpha e^{\lambda_i t}}{\alpha!} (A - \lambda_i \mathbf{I})^\alpha Z_{\lambda_i t} \ , \ t \geq 0 \ , \tag{2.5}$$

where $Z_{\lambda_i t}$ is the eigenprojection of At at $\lambda_i t$, $i = 1, ..., k$.

Remark 2.4. Formula (2.5) entails that every entry of $f(t) = e^{At}$ is a linear combination of the functions

$$e^{-\lambda_1 t}, ..., t^{\nu_1 - 1} e^{-\lambda_1 t}, ..., e^{-\lambda_k t}, ..., t^{\nu_k - 1} e^{-\lambda_k t} \ . \tag{2.6}$$

One approach has been discussed by Harris et al. [27] and Luther and Rost [39]. They showed that one obtains the sought-for eigenprojections by inverting the Wronski matrix[11] of the functions in (2.6) (evaluated at $t = 0$), which is a confluent Vandermonde matrix.[12] There were several algorithms proposed for the inversion of a Vandermonde matrix (cf. Turner [54] or Luther and Rost [39]), but Agaev and Chebotarev [4] (Proposition 2) determined a closed formula for the eigenprojections of a matrix at its eigenvalues.

Theorem 2.2. Let $A \in \mathbb{C}^{n \times n}$ be a square matrix with eigenvalues $\lambda_1, ..., \lambda_k$ and respective indexes $\nu_1, ..., \nu_k$. Let $u_1, ..., u_k$ be integers with $u_i \geq \nu_i$, $i = 1, ..., k$. If $k \geq 2$ the eigenprojection Z_{λ_i} of A at λ_i is given by

$$Z_{\lambda_i} = \frac{\prod_{j \neq i} \left((A - \lambda_i \mathbf{I})^{u_i} - (\lambda_j - \lambda_i)^{u_i} \mathbf{I} \right)^{u_j}}{\prod_{j \neq i} \left(-(\lambda_j - \lambda_i)^{u_i} \right)^{u_j}} \ . \tag{2.7}$$

If $k = 1$ we have $(A - \lambda_1 \mathbf{I})^{\nu_1} = 0$ and because of Definition 2.16 ii) $Z_{\lambda_1} = \mathbf{I}$.

[11] Named after Józef Maria Hoëne-Wronski (1778-1853), the Wronski matrix $W(t) = (W_{ij}(t))_{i,j=1,...,n}$ of $n-1$ times differentiable functions $y_1(t), ..., y_n(t)$ is given by $W_{ij}(t) = y_j^{(i-1)}(t)$, $i, j = 1, ..., n$.

[12] Named after Alexandre-Thophile Vandermonde (1735-1796), a French musician and chemist.

Remark 2.5. In the paper by Moler and van Loan [44] this approach is only presented for the case of n pairwise distinct eigenvalues $\lambda_1, ..., \lambda_n$ (the non-confluent case).

Remark 2.6. In Proposition 2 from Agaev and Chebotarev [4] the denominator is not correct. The constant term of the annihilating polynomial used in the proof is the denominator in Formula (2.7).

Remark 2.7. Formula (2.7) entails that the eigenprojection of A at λ_i equals the eigenprojection of At at $\lambda_i t$, $i = 1, ..., k$. If we calculate e^A, it is no problem to determine the function

$$f(t) := e^{At} = \sum_{i=1}^{k} \sum_{\alpha=0}^{\nu_i - 1} \frac{t^\alpha e^{\lambda_i t}}{\alpha!} (A - \lambda_i \mathbf{I})^\alpha Z_{\lambda_i} \ , \ t \geq 0 \ . \tag{2.8}$$

We will need these functions in Chapter 3.

Remark 2.8. If we implement this approach, distinguishing between the non-confluent and the confluent case is crucial. A numerical derivation of the eigenvalues often yields close but distinct eigenvalues in the confluent case. Using the non-confluent case instead will cause the denominator in (2.7) to be small, which may result in numerical problems. So we should rather assume these close eigenvalues to be equal (by rounding) and use the confluent case. For a further discussion of the properties of eigenvalue-based methods and the effects of a pertubation of the matrix (e.g. by rounding) see Moler and van Loan [44] and the references therein.

Example 2.6. We consider a matrix $Q = (q_{ij})_{i,j=1,2} \in \mathbb{R}^{2\times 2}$. The eigenvalues are the solutions of the characteristic polynomial (2.3)

$$c_Q(\lambda) = \lambda^2 - \mathrm{tr}(Q)\lambda + \det(Q) \ ,$$

given by the real numbers

$$\lambda_{1,2} = \frac{q_{11} + q_{22}}{2} \pm \sqrt{\frac{(q_{11} - q_{22})^2}{4} + q_{12} q_{21}} \ .$$

(i) The eigenvalues are equal iff $q_{12}q_{21} = 0$ and $q_{11} = q_{22} = \lambda_{1,2}$. In this case we have wlg $Q = \begin{pmatrix} q_{11} & q_{12} \\ 0 & q_{11} \end{pmatrix}$. Since $k = 1$ we obtain $Z_{q_{11}} = \mathbf{I}$ and with (2.8)

$$e^{Qt} = e^{q_{11}t}(\mathbf{I} + (Q - q_{11}\mathbf{I})t) = \begin{pmatrix} e^{q_{11}t} & q_{12}te^{q_{11}t} \\ 0 & e^{q_{11}t} \end{pmatrix} \ , \ t \geq 0 \ .$$

(ii) If $\lambda_1 \neq \lambda_2$ we have $\nu_1 = \nu_2 = 1$, and by using $u_1 = u_2 = 1$ in (2.7) we obtain

$$\begin{aligned} &Z_{\lambda_1} = (\lambda_2 - \lambda_1)^{-1}(\lambda_2 \mathbf{I} - Q) \ , \\ &Z_{\lambda_2} = (\lambda_2 - \lambda_1)^{-1}(-\lambda_1 \mathbf{I} + Q) \ , \\ &\rightarrow \ e^{Qt} = e^{\lambda_1 t}(\lambda_2 - \lambda_1)^{-1}(\lambda_2 \mathbf{I} - Q) + e^{\lambda_2 t}(\lambda_2 - \lambda_1)^{-1}(-\lambda_1 \mathbf{I} + Q) \ . \end{aligned}$$

In the special case $q_{12} = 0$ this simplifies to $\lambda_1 = q_{11}, \ \lambda_2 = q_{22}$ and

$$e^{Qt} = \begin{pmatrix} e^{q_{11}t} & (q_{11} - q_{22})^{-1} q_{12}(e^{q_{11}t} - e^{q_{22}t}) \\ 0 & e^{q_{22}t} \end{pmatrix} \ , \ t \geq 0 \ .$$

2.4 Homogeneous Markov chains

Markov[13] chains are stochastic processes which are useful in many applications like decision theory or queuing theory. The definition and the properties of Markov chains may be reviewed in many textbooks like the ones by Kijima [34], Norris [50] and Serfozo [53]. Basic properties are presented as remarks without proofs and references. We only consider a certain type of Markov chains, namely homogeneous Markov chains with a finite state space. In Chapter 3 use such processes in our model.

Definition 2.17. A right-continuous process $(J_t)_{t \in T}$ with finite state space $E = \{1, ..., n\}$ and index set $T = \mathbb{R}_0^+$ or $T = \mathbb{N}_0$ is called *homogeneous Markov chain with finite state space*, if it satisfies the Markov property

$$\begin{aligned} &\forall \ k \in \mathbb{N} \ \forall \ x_1, ..., x_k, x \in E, \ t_1, ..., t_k, t \in T : t_1 < ... < t_k < t \\ &\mathrm{P}(J_t = x | J_{t_1} = x_1, ..., J_{t_k} = x_k) = \mathrm{P}(J_{t-t_k} = x | J_0 = x_k) \ , \end{aligned} \tag{2.9}$$

whenever the probabilities are defined. A Markov chain (MC) is called

i) time-discrete if $T = \mathbb{N}_0$,

ii) time-continuous if $T = \mathbb{R}_0^+$,

The distribution of J_0 is called initial distribution of J.

In the following we investigate homogeneous Markov chains with finite state space but we omit writing *homogeneous* and *with finite state space*. We just distinguish between the time-discrete and the time-continuous case. We define the *transition probabilities*

$$p(i, j, t) := \mathrm{P}(J_t = j | J_0 = i) \ , \ i, j = 1, ..., n \ , \ t \in T$$

[13]Named after Andrey Andreyevich Markov (1856-1922), a Russian mathematician.

and the distribution of J

$$p(i,t) := \mathrm{P}(J_t = i) \ , \ i = 1, ..., n \ , \ t \in T \ .$$

For a time-discrete MC we define the one-step transition probabilities π_{ij} for a jump from state i to state $j \neq i$ by

$$\pi_{ij} = p(i,j,1) \ , \ i,j = 1, ..., n$$

and the probability π_{ii} that the MC stays in state i at time t by

$$\pi_{ii} = p(i,i,1) \ , \ i = 1, ..., n.$$

The matrix $\Pi := (\pi_{ij})_{i,j=1,...,n}$ is called transition matrix. Π is a stochastic matrix ($\Pi \geq 0$ and $\Pi \mathbf{1} = \mathbf{1}$). The probabilities $p(i,j,t) \ , \ i,j = 1, ..., n \ , \ t \in \mathbb{N}$ may be calculated recursively

$$p(i,j,t) = \sum_{k=1}^{n} p(i,k,t-1)p(k,j,1) \quad , \ i,j = 1, ..., n \ ,^{14}$$

and hence the corresponding t-step transition matrix $\Pi^{(t)} = (p(i,j,t))_{i,j=1,...,n}$ equals $\Pi^t \ , \ t \in \mathbb{N}$. So the distribution of a time-discrete MC is completely described by the transition matrix Π and the initial distribution a, where $a = (a_i)_{i=1,...,n}$ with $a_i := p(i,0) \ , \ i = 1, ..., n$. For a time-discrete MC we call the pair (a, Π) representation of the MC.
For a time-continuous MC we also need the transition probabilities $p(i,j,t)$, $t \geq 0, \ i,j = 1, ..., n$. Since the paths of a MC are right-continuous we have

$$\lim_{t \to 0+} p(i,i,t) = p(i,i,0) = 1 \ , i = 1, ..., n \ .$$

This ensures that for $i = 1, ..., n$ there exists $q_i \geq 0$ so that

$$q_i = \lim_{t \to 0+} t^{-1}(1 - p(i,i,t)) \ , \ i = 1, ..., n \ .$$

The quantity q_i is called exit rate of the state i. Further we define the quantities $q_{ij} \ , \ i,j = 1, ..., n, \ i \neq j$ by

$$q_{ij} = \lim_{t \to 0+} t^{-1} p(i,j,t) \ , \ i,j = 1, ..., n, \ i \neq j \ .$$

These quantities are called jump-rates, q_{ij} is the rate of a jump from state i to state j. Obviously $q_i = \sum_{i \neq j} q_{ij} \ , \ i = 1, ..., n$ holds. With $q_{ii} := -q_i$ for $i = 1, ..., n$ the matrix

$$Q = (q_{ij})_{i,j=1,...,n}$$

is called intensity matrix of the MC.

[14] This Formula is a discrete version of the Chapman-Kolmogorov equality (2.11).

Remark 2.9. Let $Q = (q_{ij})_{i,j=1,...,n}$ be the intensity matrix of a time-continuous MC J. Then

i) $q_{ij} \geq 0$ for $i \neq j$, $q_{ii} \leq 0$ for $i = 1, ..., n$, $\sum_{j=1}^{n} q_{ij} = 0$ for $i = 1, ..., n$,

ii) for $i, j = 1, ..., n$ the transition probabilities satisfy the forward equation

$$\frac{\partial}{\partial t} p(i, j, t) = \sum_{k=1}^{n} p(i, k, t) q_{kj} \ , \ t \geq 0 \ ,$$

whose solution is given by

$$p(i, j, t) = e_i' e^{Qt} e_j \ , \quad i, j = 1, ..., n \ ,$$

where e^{Qt} is the matrix exponential of Qt.

iii) the one-step transition matrix Π defined by

$$\begin{aligned} \pi_{ij} &= -\frac{q_{ij}}{q_{ii}} \quad , \ i \neq j \ , \pi_{ii} = 0 \ , \text{ if } q_{ii} \neq 0, \\ \pi_{ij} &= 0 \qquad\quad , \ i \neq j \ , \pi_{ii} = 1 \ , \text{ if } q_{ii} = 0. \end{aligned} \tag{2.10}$$

gives the conditional distribution of J

$$\mathrm{P}(J_\tau = j | J_0 = i) = \pi_{ij} \ , \ i, j = 1, ..., n \ , \ q_{ii} \neq 0 \ ,$$

at the first jump-time $\tau = \inf\{t \geq 0 : J_t \neq J_0\}$.

The distribution of a time-continuous MC J is completely described by the intensity matrix Q and the initial distribution a. We call the pair (a, Q) representation of the time-continuous MC J.

Remark 2.10. For $i = 1, ..., n$ with $q_{ii} \neq 0$ the first jump time τ is conditional exponentially distributed[15]:

$$\mathrm{P}^{\tau | J_0 = i} = \mathrm{Exp}(-q_{ii}) \ .$$

Remark 2.11. For every time-continuous finite homogeneous MC J with one-step transition matrix Π the process $\tilde{J} = (J_{\tau_i})_{i \in \mathbb{N}_0}$, where $\tau_0 = 0$ and $(\tau_i)_{i \in \mathbb{N}}$ are the jump times of J, is a time-discrete MC with transition matrix Π from (2.10). If we order the states of J such that $q_{ii} < 0$ for $i = 1, .., m$ and $q_{ii} = 0$ for $i = m + 1, ..., n$ we may write

$$Q = \begin{pmatrix} Q_m & Q_0 \\ 0 & 0 \end{pmatrix},$$

[15]Which means $\mathrm{P}(\tau > t | J_0 = i) = e^{q_{ii} t}$, $t \geq 0$.

where Q_m is an $m \times m$ matrix. The matrix Π may now be written as

$$\Pi = \begin{pmatrix} \Pi_m & \Pi_0 \\ 0 & \mathbf{I} \end{pmatrix},$$

where

$$\Pi_m = \mathbf{I} - \text{diag}(q_{11}^{-1}, ..., q_{mm}^{-1}) \cdot Q_m \text{ and } \Pi_0 = -\text{diag}(q_{11}^{-1}, ..., q_{mm}^{-1}) \cdot Q_0.$$

So we may calculate Q if we know Π and the exit intensities $q_{11}, ..., q_{mm}$:

$$Q = \begin{pmatrix} \text{diag}(q_{11}, ..., q_{mm}) \cdot (\mathbf{I} - \Pi_m) & -\text{diag}(q_{11}, ..., q_{mm}) \cdot \Pi_0 \\ 0 & 0 \end{pmatrix}.$$

Remark 2.12. The matrices $\left(e^{Qt}\right)_{t \geq 0}$ ($(\Pi^n)_{n \in \mathbb{N}_0}$ in the time-discrete case) are linear operators with the semigroup property[16]

$$e^{Qt} e^{Qs} = e^{Q(t+s)} \ , \ t, s \geq 0 \ , \quad (\text{resp. } \Pi^n \Pi^m = \Pi^{n+m} \ , \ n, m \in \mathbb{N}_0) \ , \qquad (2.11)$$

that is also known as *Chapman-Kolmogorov equality*[17][18].

The next theorem which is proved by e.g. Norris [50] (Theorem 1.4.2 and Theorem 6.5.4) generalizes the Markov property (2.9).

Theorem 2.3. Strong Markov property:
Let $(J_t)_{t \in T}$ be a MC with representation (a, Q) in the time-continuous case with $T = \mathbb{R}_0^+$ (resp. (a, Π) in the time-discrete case with $T = \mathbb{N}_0$) and τ a stopping time with respect to the canonical filtration $\mathfrak{F}^J$. Then, conditional on $\tau < \infty$ and $J_\tau = i$, $(J_{\tau+t})_{t \in T}$ is a MC with representation (e_i, Q) (resp. (e_i, Π)) and independent of $(J_s)_{s \leq \tau}$.

We now investigate absorbing MCs.

Definition 2.18. A state i of a MC J is called

i) *recurrent* if

$$\lim_{t \to \infty} p(i, i, t) > 0 \ ,$$

[16] A semigroup is a set S together with a binary operation $\times$ that satisfies:
- Closure: For all $a, b \in S$, the result of the operation $a \times b$ is also in S.
- Associativity: For all $a, b, c \in S$, the equation $(a \times b) \times c = a \times (b \times c)$ holds.

[17] Named after Sydney Chapman (1888-1970), a British mathematician and geophysicist and Andrey Nikolaevich Kolmogorov (1903-1987), a Soviet Russian mathematician.

[18] In general we have for all commutating matrices A, B: $e^A e^B = e^{A+B}$.

ii) *transient* if

$$\lim_{t\to\infty} p(i,i,t) = 0 \ ,$$

iii) *absorbing* if

$$p(i,i,t) = 1 \ , \ \forall \ t \in T \ .$$

A MC J is called absorbing if every state i is either transient or absorbing. Two states i, j of a MC J are *connected* if

$$\exists \ t_1, t_2 \geq 0 : \ p(i,j,t_1) > 0 \ \ , \ \ p(j,i,t_2) > 0 \ .$$

A MC which states are all connected is called *irreducible*.

If a state i is transient, then also

$$\lim_{t\to\infty} p(j,i,t) = 0 \ , \ \forall \ j = 1, ..., n \ .$$

Consequently,

$$\lim_{t\to\infty} p(i,t) = \sum_{j=1}^{n} a_j p(j,i,t) = 0 \ .$$

In the notation of Remark 2.11 the absorbing states are the states $m+1, ..., n$. Now we proof three lemmas on absorbing Markov chains where we use the notation of Remark 2.11.

Lemma 2.2. A time-continuous MC J is absorbing iff the matrix Q_m that corresponds to the non-absorbing states is regular. If J is absorbing, let $\tau := \inf\{t \geq 0 : J_t > m\}$ be the time to absorption. The probabilities

$$b_{ij} := \mathrm{P}(J_\tau = j + m | J_0 = i) \ , \ i = 1, ..., m, \ j = 1, ..., n - m \ ,$$

that, starting in state i, the chain is absorbed in state $j + m$, are given by the entries of the $m \times (n - m)$-matrix

$$B = -Q_m^{-1} Q_0 \ .$$

The first part is proven by Neuts [47] (Lemma 2.2.1). The absorption probabilities are also derived by Kijima [34] (Section 4.6). We have more general assumptions than Kijima[19], but his proof is still correct here. We present a different approach:

Proof: We use Theorem 3.3.1. from Norris [50] which states that for every

[19] Kijima investigates *lossy generators* which have the property that Q_m is irreducible.

$j = m+1, ..., n$ the probabilities $b_{ij} := \mathrm{P}(J_\tau = j+m | J_0 = i)$, $i = 1, ..., n$ solve the system of linear equations

$$\begin{aligned} b_{ij} &= 1 \qquad\qquad , \text{ for } i = j \ , \\ \sum_{k=1}^{n} q_{ik} b_{kj} &= 0 \quad , \text{ for } i \neq j \ . \end{aligned}$$

Since $b_{kj} = 1$ for $k = j$ and $b_{kj} = 0$ for $j = m+1, ..., n$, $k \neq j$ this simplifies to

$$\sum_{k=1}^{m} q_{ik} b_{kj} + q_{ij} = 0 \ , \ i = 1, ..., m \ .$$

We may write this in matrix form

$$Q_m B + Q_0 = 0 \ ,$$

and because Q_m is regular this equation has the unique solution

$$B = -Q_m^{-1} Q_0 \ .$$

□

Lemma 2.3. Let J be an absorbing MC with transient states $\{1, ..., m\}$ and $\tau = \inf\{t \geq 0 : J_t > m\}$ the time to absorption. Then the mean time to absorption is given by

$$\mathrm{E}(\tau) = -a' Q_m^{-1} \mathbf{1} \ .$$

Proof: Since every state $1, ..., m$ is transient, the values

$$\mu_i^\tau := \mathrm{E}(\tau | J_0 = i) \ , \ i = 1, ..., m \ ,$$

are finite and by conditioning on the first jump $\xi := \inf\{t \geq 0 : J_t \neq J_0\}$ of J we obtain with Theorem 2.3

$$\begin{aligned} \mu_i^\tau &= \sum_{i=1}^{n} \mathrm{E}(\tau | J_0 = i, J_\xi = j) \mathrm{P}(J_\xi = j | J_0 = i) \\ &= \frac{1}{-q_{ii}} + \sum_{j=1, j \neq i}^{m} \frac{q_{ij}}{-q_{ii}} \mu_j^\tau \quad , \ i = 1, ..., m \ . \end{aligned}$$

This simplifies to

$$\sum_{j=1}^{m} q_{ij} \mu_j^\tau = -1 \quad , \ i = 1, ..., m \ ,$$

and if we use $\mu^\tau = (\mu_1^\tau, ..., \mu_m^\tau)'$ we have

$$Q_m \mu^\tau = -\mathbf{1} \;\Rightarrow\; \mu^\tau = -Q_m^{-1}\mathbf{1} \;\Rightarrow \mathrm{E}(\tau) = a'\mu^\tau = -a'Q_m^{-1}\mathbf{1}$$

because Q_m is regular (Lemma 2.2). □
We will have a different derivation of this result in Lemma 2.6.

Lemma 2.4. Let J be an absorbing MC with transient states $\{1, ..., m\}$, $\tau = \inf\{t \geq 0 : J_t > m\}$ the time to absorption and B be the matrix of absorption probabilities from Lemma 2.2. Let

$$\mu_{ij} := \begin{cases} \mathrm{E}(\tau | J_0 = i, J_\tau = j + m) & , \; b_{ij} > 0 \\ \infty & , \; b_{ij} = 0 \end{cases} ,$$

be the mean time to absorption if J starts in state i and is absorbed into state j, $i = 1, ..., m$, $j = 1, ..., n - m$. Then

$$\mu_{ij} = \begin{cases} \dfrac{d_{ij}}{b_{ij}} & , \; b_{ij} > 0 \\ \infty & , \; b_{ij} = 0 \end{cases} , \; i = 1, ..., m, \; j = 1, ..., n - m \; ,$$

where $D = (d_{ij})_{i=1,...,m, \; j=1,...,n-m}$ is given by

$$D = -Q_m^{-1}B = Q_m^{-2}Q_0 \; .$$

Proof: Let $0 \cdot \infty := 0$ to have all terms well-defined. Let

$$\xi := \inf\{t \geq 0 : J_t \neq J_0\}$$

be the first jump-time of J. Now for every $i = 1, ..., m$, $j = 1, ..., n - m$ with $b_{ij} > 0$ we may condition on J_ξ to obtain

$$\begin{aligned} \mu_{ij} &= \mathrm{E}(\tau | J_0 = i, J_\tau = j + m) \\ &= \sum_{k=1}^{n} \mathrm{E}(\tau | J_0 = i, J_\tau = j + m, J_\xi = k) \mathrm{P}(J_\xi = k | J_0 = i, J_\tau = j + m) \; . \end{aligned}$$

The terms in the sum with $k = i$ or $m < k \neq j + m$ vanish since for these summands $\mathrm{P}(J_\xi = k | J_0 = i, J_\tau = j + m) = 0$. For the remaining terms we have

$$\begin{aligned} \mathrm{P}(J_\xi = k | J_0 = i, J_\tau = j + m) &= \frac{\mathrm{P}(J_\xi = k, J_0 = i, J_\tau = j + m)}{\mathrm{P}(J_0 = i, J_\tau = j + m)} \\ = \frac{\mathrm{P}(J_\tau = j + m | J_0 = i, J_\xi = k) \mathrm{P}(J_\xi = k | J_0 = i)}{\mathrm{P}(J_\tau = j + m | J_0 = i)} &= \begin{cases} \dfrac{q_{ik}}{-q_{ii} b_{ij}} & , \; k = j + m \\ \dfrac{q_{ik} b_{kj}}{-q_{ii} b_{ij}} & , \; k \neq j + m \end{cases} , \end{aligned}$$

and

$$\mathrm{E}(\tau|J_0 = i, J_\tau = j+m, J_\xi = k) = \begin{cases} -q_{ii}^{-1} & , \ k = j+m \\ -q_{ii}^{-1} + \mu_{kj} & , \ k \neq j+m \ . \end{cases}$$

We obtain the linear system of equations

$$-q_{ii}\mu_{ij}b_{ij} = \frac{1}{-q_{ii}} \left(q_{i,j+m} + \sum_{k=1,k\neq i}^{m} q_{ik}b_{kj} \right) + \sum_{k=1,k\neq i}^{m} q_{ik}b_{kj}\mu_{kj} \ ,$$

for $i = 1, ..., m$, $j = 1, ..., n-m$. From Lemma 2.2 we know

$$\sum_{k=1}^{m} q_{ik}b_{kj} = -q_{i,j+m} \ ,$$

which leads to

$$\sum_{k=1}^{m} q_{ik}b_{kj}\mu_{kj} = -b_{ij} \ , \ i = 1, ..., m, \ j = 1, ..., n-m \ . \tag{2.12}$$

If we define the matrix $D = (d_{ij})_{i=1,...,m, \ j=1,...,n-m}$ by $d_{ij} = b_{ij}\mu_{ij}$, we may solve (2.12) resulting in

$$Q_m D = -B \ \Rightarrow \ D = -Q_m^{-1}B \ .$$

□

Remark 2.13. These three lemmas may be generalized for any subset of transient states of a not necessarily absorbing time-continuous MC: Let J be a time-continuous MC on $\{1, ..., n\}$ with parameters $a, Q = \begin{pmatrix} Q_m & Q_0 \\ A & B \end{pmatrix}$. If wlg the states $\{1, ..., m\}$ are transient let $\tau := \inf\{t \geq 0 : J_t > m\}$ be the first time when J is not in $\{1, ..., m\}$. Now the stopped MC $\tilde{J}$ with $\tilde{J}_t = J_{t\wedge\tau}$, $t \geq 0$ is absorbing with parameters $a, \tilde{Q}$ where

$$\tilde{Q} = \begin{pmatrix} Q_m & Q_0 \\ 0 & 0 \end{pmatrix} \ ,$$

and $\tau = \inf\{t \geq 0 : \tilde{J}_t > m\}$, $\mathrm{P}(J_\tau = j) = \mathrm{P}(\tilde{J}_\tau = j)$, $j = m+1, ..., n$. This also yields $\mathrm{E}(\tau|J_0 = i, J_\tau = j) = \mathrm{E}(\tau|\tilde{J}_0 = i, \tilde{J}_\tau = j)$, $i = 1, ..., m$, $j = m+1, ..., n$.

Next we investigate the long time behavior of a time-continuous MC. A probability vector $x = (x_1, ..., x_n)'$ is called *steady-state distribution* of a Markov chain with intensity matrix Q, if

$$x'Q = 0 \ , \quad x'\mathbf{1} = 1 \ , \quad x_i \geq 0 \ , \ i = 1, ..., n \ .$$

For every initial distribution a and intensity matrix Q the vector x^* given by

$$x_i^* = \lim_{t \to \infty} p(i,t) \ , \ i = 1, ..., n$$

is a steady-state distribution but, in general, x^* is not the only steady state distribution. We may renumber the states of the MC in a way that the intensity matrix Q has the form

$$Q = \begin{pmatrix} Q_{00} & Q_{01} & Q_{02} & \cdots & Q_{0k} \\ 0 & Q_1 & 0 & \cdots & 0 \\ \vdots & \ddots & Q_2 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & Q_k \end{pmatrix} \ , \tag{2.13}$$

where $Q_{00} \in \mathbb{R}^{n_0 \times n_0}$, $Q_l \in \mathbb{R}^{n_l \times n_l}$, $l = 0, ..., k$, the states $1, ..., n_0$ are transient and for every $l \in \{1, ..., k\}$ the states in block $B_l = \{n_0 + ... + n_{l-1} + 1, ..., n_0 + ... + n_l\}$ are connected and recurrent (the blocks are then called irreducible). For every $l = 1, ..., k$ the matrix Q_l is an intensity matrix of an irreducible MC. An irreducible time-continuous MC has an unique steady-state distribution.

Lemma 2.5. Let Q be the intensity matrix of a MC J like in (2.13) with transient states $1, ..., n_0$ and irreducible blocks

$$B_l = \{n_0 + ... + n_{l-1} + 1, ..., n_0 + ... + n_l\} \ , \ l = 1, ..., k \ .$$

The probability z_{ul} of entering block B_l when starting in the transient state u is

$$z_{ul} = -e_u Q_{00}^{-1} Q_{0l} \mathbf{1} \ , \ u = 1, .., n_0 \ , \ l = 1, ..., k \ .$$

The steady-state distribution of J in dependence of the initial distribution $a = (\mathbf{a}_0, ..., \mathbf{a}_k)'$ ($\mathbf{a}_l$ is a vector of length n_l, $l = 0, ..., k$) is given by:

$$x^a = \begin{pmatrix} 0 \\ (\mathbf{a}_1'\mathbf{1} + \mathbf{a}_0' z_{\cdot 1}) y^1 \\ \vdots \\ (\mathbf{a}_k'\mathbf{1} + \mathbf{a}_0' z_{\cdot k}) y^k \end{pmatrix} \ ,$$

where y^l is the unique probability vector that solves

$$(y^l)'Q_l = \mathbf{0} \ , \ l = 1, ..., k \ .$$

Proof: We already know that $p(i,t) \to 0$ for a transient state $i = 1, ..., n_0$. Let $\tau := \inf\{t \geq 0 : J_t > n_0\}$ be the first time when J is not in a transient state. Because of the strong Markov property from Theorem 2.3

$$\lim_{t\to\infty} \mathrm{P}(J_{t+\tau} = n_0 + ... + n_{l-1} + j | J_\tau \in B_l) = y_j^l \ , \ j = 1, ..., n_l \ , \ l = 1, ..., k \ .$$

Next

$$\mathrm{P}(J_\tau \in B_l) = \mathrm{P}(J_0 \in B_l) + \sum_{u=1}^{n_0} \mathrm{P}(J_\tau \in B_l | J_0 = u)\mathrm{P}(J_0 = u) \ , \ l = 1, ..., k \ .$$

Clearly $\mathrm{P}(J_0 \in B_l) = \mathbf{a}_l'\mathbf{1} \ , \ l = 1, ..., k$ and for $z_{ul} = \mathrm{P}(J_\tau \in B_l | J_0 = u)$ we again use Theorem 3.3.1. from Norris [50] to obtain for every $l = 1, ..., k$ the system of linear equations

$$\sum_{v=1}^{n_0} q_{uv} z_{vl} + \sum_{w=1}^{n_l} q_{u,n_0+...+n_{l-1}+w} = 0 \ , \ u = 1, ..., n_0 \ .$$

We may write in matrix form

$$Q_{00}Z + \overline{Q} = 0 \ ,$$

where $\overline{Q}$ is an $n_0 \times k$ matrix which columns are given by $Q_{01}\mathbf{1}, ..., Q_{0k}\mathbf{1}$. Since Q_{00} is regular the unique solution is

$$Z = -Q_{00}^{-1}\overline{Q} \ .$$

All together if $i = j + n_0 + ... + n_{l-1} \in B_l$ for $t \to \infty$

$$p(i, t+\tau) = \mathrm{P}(J_{t+\tau} = i | J_\tau \in B_l)\mathrm{P}(J_\tau \in B_l) \to y_j^l(\mathbf{a}_l'\mathbf{1} + \mathbf{a}_0' z_{\cdot l}) \ ,$$

and since τ is finite P-a.s. this proves the Lemma. □

Remark 2.14. For $l = 1, ..., k$ the unique probability vector that solves $(y^l)'Q_l = 0$ is the unique solution of

$$(y^l)'Q_l = 0 \text{ and } (y^l)'\mathbf{1} = 1 \ . \tag{2.14}$$

The vector $\mathbf{1}$ is independent from the columns of Q_l and every $n_l - 1$ columns of $Q_l \in \mathbb{R}^{n_l \times n_l}$ are independent. We define the matrix $\tilde{Q}_l \in \mathbb{R}^{n_l \times n_l}$ by taking Q_l and replacing the last column with the vector $\mathbf{1}$. The matrix $\tilde{Q}_l$ is regular and the solution of (2.14) is given by

$$(y^l)' = e_{n_l}' \tilde{Q}_l^{-1} \ .$$

2.5 Phase-type distributions

In this section we introduce the class of phase-type distributions, which was first described by Neuts [45]. The concept of phases is closely related to the model introduced in Chapter 3.
Let J be a time-continuous homogeneous MC with state space $\{1, ..., m+1\}$, where the states $1, ..., m$ are transient and the state $m+1$ is absorbing. The intensity matrix Q equals analogous to Remark 2.11

$$Q = \begin{pmatrix} Q_m & Q_0 \\ 0 & 0 \end{pmatrix},$$

where Q_m is regular. We assume that the initial distribution is concentrated on the transient states and define $a = (\mathrm{P}(J_0 = 1), ..., \mathrm{P}(J_0 = m))'$.

Definition 2.19. The distribution of the time to absorption of the MC J

$$\tau := \inf \{t > 0 : J_t = m + 1\} \tag{2.15}$$

is called *phase-type* distribution with representation (a, Q_m) and we write short $\tau \sim \mathrm{PH}(a, Q_m))$.

The matrix Q_m also yields the vector Q_0 because $Q_0 = -Q_m\mathbf{1}$. The representation is not unique. For any pair of parameters a, Q_m there exist infinitely many pairs of parameters $\tilde{a}, \tilde{Q}_{\tilde{m}}$ with $\mathrm{PH}(a, Q_m) = \mathrm{PH}(\tilde{a}, \tilde{Q}_m)$ (cf. Neuts [47]).

Lemma 2.6. Let τ be phase-type distributed with representation (a, Q_m). Then τ has the

i) cumulative distribution function

$$F(t) = \begin{cases} 1 - a'e^{tQ_m}\mathbf{1} & , \ t \geq 0 \\ 0 & , \ t < 0 \ , \end{cases}$$

ii) mean

$$\mathrm{E}(\tau) = -a'Q_m^{-1}\mathbf{1} \ ,$$

iii) density function

$$f(t) = \begin{cases} a'e^{tQ_m}Q_0 & , \ t \geq 0 \ , \\ 0 & , \ t < 0 \ , \end{cases}$$

The Lemma was proven by Neuts [45] and we recognize our result for the mean from Lemma 2.3. The class of phase-type distributions includes roughly spoken all finite mixtures of convolutions of exponential distributions, in particular the exponential, the hyper-exponential and the Erlang distribution. For further pros of these distributions we quote in extracts Ahn and Ramaswami [5] Section 1.1:

> "PH-distributions have received much attention in the applied probability literature related to queues, dams, insurance risks, reliability, etc., and the reasons for that have been many:
>
> – Denseness: The class of PH-distributions can be shown to be dense (...) in the set of all probability distributions on $[0, \infty)$. (...)
> – Closure: The class of PH-distributions is closed under finite convolutions and mixtures and under Boolean operations of taking the max or min of (independent) PH-random variables. (...)
> – Tractability: A very attractive feature of phase type distributions is their computational tractability. Due to the connection with an underlying Markov chain, in models involving phase type distributions, conditioning arguments become easier through the inclusion of the state of the Markov chain as an auxiliary variable."

The way how phase-type distributions are closed under convolutions is specified in the following Lemma.

Lemma 2.7. Let $X_1 \sim \mathrm{PH}(a, Q_{m_1})$ and $X_2 \sim \mathrm{PH}(b, R_{m_2})$ be independent. Then the sum $X_1 + X_2$ is again phase-type distributed with representation

$$\begin{pmatrix} a \\ 0 \end{pmatrix}, \begin{pmatrix} Q_{m_1} & Q_0 b' \\ 0 & R_{m_2} \end{pmatrix}.$$

A proof of this Lemma is given by Neuts [47] (Theorem 2.2.2.). The benefit of this Lemma is that we may obtain the convolution of PH-distributions by using matrix calculus instead of numerical integration.

2.6 Martingales and compensator

In this section we present the framework of decomposing counting processes. Pioneer work was the famous Doob-Meyer decomposition, see Doob [23] and Meyer [41, 42]. The theoretical background is conveniently presented in Appendix A in Aven and Jensen [8]. We only highlight the results needed in our further investigations.

Definition 2.20. Let X be a $\mathbb{R}$-valued random variable on $(\Omega, \mathfrak{A}, \mathrm{P})$ and let $\mathfrak{A}_0$ be a sub-σ-algebra $\mathfrak{A}_0 \subseteq \mathfrak{A}$. A $\mathfrak{A}_0$-measurable $\mathbb{R}$-valued random variable Z is called (a version of) the *conditional expectation* of X given $\mathfrak{A}_0$ if

$$\int_A X \mathrm{dP} = \int_A Z \mathrm{dP} \ , \ \forall \ A \in \mathfrak{A}_0$$

and we write $Z = \mathrm{E}(X|\mathfrak{A}_0)$.

If $\mathrm{E}|X|^2 < \infty$ the conditional expectation $\mathrm{E}(X|\mathfrak{A}_0)$ is the projection (regarding the norm $\| X \| = \mathrm{E}|X|^2$) of X onto the subspace of $\mathfrak{A}_0$-measurable random variables Y with $\mathrm{E}|Y|^2 < \infty$. Especially, $\mathrm{E}(X|\mathfrak{A}_0) = X$ if X is $\mathfrak{A}_0$-measurable and $\mathrm{E}(X|\mathfrak{A}_0) = \mathrm{E}(X)$ if X is independent of $\mathfrak{A}_0$ (we use this for $\mathfrak{A}_0 = \{\emptyset, \Omega\}$, since X is always independent of $\{\emptyset, \Omega\}$). Two versions of the conditional expectation are equal P-a.s. If there exists an E-valued random variable Y with $\mathfrak{A}_0 = \sigma(Y)$, because of Lemma 2.1 there exists a mapping $h : E \to \mathbb{R}$ with

$$\mathrm{E}(X|\mathfrak{A}_0) = h(Y) \ . \tag{2.16}$$

Definition 2.21. Let $X = (X_t)_{t \in T}$ be a $\mathbb{R}$-valued stochastic process adapted to the filtration $\mathfrak{F} = (\mathfrak{F}_t)_{t \in T}$.

i) X is called *integrable* if $\mathrm{E}|X_t| < \infty \ \forall \ t \in T$.

ii) X is called $\mathfrak{F}$-*progressive* if $\forall \ t \in T$ the mapping $(s, \omega) \to X_s(\omega)$ on $[0, t] \times \Omega$ is measurable with respect to the σ-algebra $\mathfrak{B}([0, t]) \otimes \mathfrak{F}_t$.

iii) X is called $\mathfrak{F}$-*predictable* if the mapping $(t, \omega) \to X_t(\omega)$ on $(0, \infty) \times \Omega$ into $\mathbb{R}$ is $\mathcal{P}(\mathfrak{F})$-measurable, where the σ-algebra $\mathcal{P}(\mathfrak{F})$ is generated by the sets

$$(s, t] \times A, \ 0 \leq s < t, \ A \in \mathfrak{F}_s, \ t \geq 0 \ .$$

iv) X is called *martingale* if $\mathrm{E}(X_t|\mathfrak{F}_s) = X_s \ \forall \ s \leq t \in T$. We denote by $\mathcal{M}_0(\mathfrak{F})$ the set of martingales with cadlag paths and with $M_0 = 0$ P-a.s.

Every left- or right-continuous adapted process is $\mathfrak{F}$-progressive. Every left-continuous adapted process is $\mathfrak{F}$-predictable and every $\mathfrak{F}$-predictable process is $\mathfrak{F}$-progressive. If X is $\mathfrak{F}$-predictable, then X_t is $\mathfrak{F}_{t-}$-measurable $\forall \ t$.

Theorem 2.4. Let $N = (N_t)_{t \geq 0}$ be an integrable counting process adapted to the filtration $\mathfrak{F}$. Then there exists a unique $\mathfrak{F}$-predictable process $A = (A_t)_{t \geq 0}$ such that $M := N - A \in \mathcal{M}_0(\mathfrak{F})$

This is Theorem 10 from Aven and Jensen [8]. A is called the $\mathfrak{F}$-*compensator* of N. Let now N be adapted to $\mathfrak{F}$. In our applications there always exists a non-negative $\mathfrak{F}$-progressive process $\overline{\lambda} = (\overline{\lambda}_t)_{t\geq 0}$ such that

$$A_t = \int_0^t \overline{\lambda}_s \mathrm{d}s \ \ , \ t \geq 0 \ . \tag{2.17}$$

The process $\overline{\lambda}$ is called $\mathfrak{F}$-*intensity* of N and the pair $(\overline{\lambda}, M)$ is called $\mathfrak{F}$-*SSM representation* of N (smooth semimartingale representation).

Remark 2.15. If N has the SSM representation $(\overline{\lambda}, M)$, the *mean value function* $\mu(t) := \mathrm{E}(N_t)$ is given by

$$\mu(t) = \mathrm{E}(N_t) = \mathrm{E}(A_t + M_t) = \int_0^t \mathrm{E}(\overline{\lambda}_u)\mathrm{d}u \ , \ t \geq 0 \ .$$

Theorem 2.5. If N has the $\mathfrak{F}$-intensity $\overline{\lambda}$ we may calculate $\overline{\lambda}_t$ by

$$\overline{\lambda}_t = \lim_{h\to 0+} h^{-1}\mathrm{E}(N_{t+h} - N_t|\mathfrak{F}_t) = \lim_{h\to 0+} h^{-1}\mathrm{P}(N_{t+h} - N_t = 1|\mathfrak{F}_t) \ \ , \ t \geq 0 \ . \tag{2.18}$$

So, $\overline{\lambda}_t$ is the conditional intensity for a jump of N at time t.

Theorem 2.6. Let N have the $\mathfrak{F}$-intensity $\overline{\lambda}$ and $\mathfrak{G}$ be a subfiltration of $\mathfrak{F}$ such that N is adapted to $\mathfrak{G}$. Then N has also a $\mathfrak{G}$-intensity $\overline{\nu}$ and

$$\overline{\nu}_t = \mathrm{E}(\overline{\lambda}_t|\mathfrak{G}_t) \ \ , \ t \geq 0 \ .$$

This theorem follows from Theorem 12 (projection theorem) from Aven and Jensen [8] with $\mathrm{E}(N_t|\mathfrak{G}_t) = N_t$.

2.7 Alias algorithm

We want to generate random variables X from a probability distribution Q on a finite set $\Omega \subset \mathbb{R}$. The standard approach is to generate a uniformly distributed random variable $U \sim \mathrm{U}([0,1])$, and calculate the (generalized) inverse

$$F_X^{-1}(x) := \inf\{y \in \mathbb{R} : F_X(y) \geq x\} \ \ , \ 0 \leq x \leq 1$$

of the cdf F_X. Then $F_X^{-1}(U)$ and X have the same distribution (cf. Devroye [22], also for alternative methods). The alias algorithm is another approach. It was developed by Walker [55] and refined by Kronmal and Peterson [36]. The latter proved this theorem:

Theorem 2.7. Any probability measure Q on a finite set $\Omega = \{s_1, ..., s_n\}$ can be expressed as an equiprobable mixture of n distributions $\mathrm{Q}_1, ..., \mathrm{Q}_n$ on $\{s_1, ..., s_n\}$

$$\mathrm{Q}(A) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Q}_i(A) \ , \ A \subseteq \{s_1, ..., s_n\} \ ,$$

such that there exist $a_1, ..., a_n \in \{s_1, ..., s_n\}$ with

$$0 < k_i := \mathrm{Q}_i(s_i) = 1 - \mathrm{Q}_i(a_i) \ , \ i = 1, ..., n \ .$$

The proof of this theorem is constructive, and the distributions $\mathrm{Q}_1, ..., \mathrm{Q}_n$ are easily computed. We call k_i the cutoff value for s_i, the respective other mass point a_i of Q_i is called alias of s_i.
The alias algorithm needs a generator for uniform-distributed random numbers, which is usually implemented in standard statistic software. The algorithm is given by the following four steps:

1. Generate a uniformly distributed random variable $U \sim \mathrm{U}([0, n])$.
2. Set $I := \lfloor U \rfloor + 1$, now $I \sim \mathrm{U}(\{1, ..., n\})$.
3. Set $V := U - I$, now $V \sim \mathrm{U}([0, 1])$ and V and I are independent.
4. If $V \leq k_I$ return s_I, else return a_I.

The probability for obtaining the value s_i equals $\mathrm{Q}(s_i)$, $i = 1, ..., n$. The benefit of this algorithm is that we only need one comparison of values instead of up to $n - 1$ comparisons if we use the inverse of the cdf.

Chapter 3

Failure model

In this chapter we define the failure model that is examined in this thesis and we develop important analytic results about the behavior of the processes that we consider in this model. We use PH-distributed times to failure, and we allow different types of failures which are represented by failure states, so minor and major failures may be modeled. We also have a closer look on the obtained failure rates.

3.1 The time to failure

For the developing of the condition of a system we consider a homogeneous Markov chain $J = (J_t)_{t\geq 0}$ with state space $\{1, ..., n\}$ $(n \geq 2)$, initial distribution $\tilde{a}$ and intensity matrix Q such that

i) the states $1, ..., m$ with $m < n$ are transient (working states):

$$\lim_{t\to\infty} \mathrm{P}(J_t \leq m) = 0 \ , \tag{3.1}$$

ii) the states $m + 1, ..., n$ are absorbing (failure states):

$$q_{ii} = 0 \quad , i = m + 1, ..., n \ , \tag{3.2}$$

iii) the system starts in a working state:

$$\mathrm{P}(J_0 > m) = 0 \ . \tag{3.3}$$

Within this setup the random times spent in state i are exponentially distributed with intensity $-q_{ii}$ for $i = 1, ..., m$ and the probability for a jump

into state j being in state i equals $-q_{ij}/q_{ii}$ for $i = 1, ..., m$, $j = 1, ..., n$. The time to failure of the system is

$$\tau = \inf\{t \geq 0 : J_t > m\}. \tag{3.4}$$

We define $a = (\tilde{a}_1, ..., \tilde{a}_m)'$ (since (3.3) $\tilde{a}_{m+1} = ... = \tilde{a}_n = 0$) and

$$Q = \begin{pmatrix} Q_m & Q_0 \\ 0 & 0 \end{pmatrix}$$

where $Q_m = (q_{ij})_{i,j=1,...,m}$ is the sub-intensity matrix belonging to the working states. The distribution of τ is given by

$$\mathrm{P}(\tau > t) = \mathrm{P}(J_t \leq m) = \sum_{i,j=1}^{m} a_i \mathrm{P}(J_t = j | J_0 = i) = a' e^{tQ_m} \mathbf{1} \ , \tag{3.5}$$

which is a PH distribution defined in (2.15). The difference to the definition of Neuts [45] is that we consider various absorbing states. We may interpret the absorbing states as distinct types of failure. We usually omit writing the rows of Q belonging to the failure states, since all entries are zero.

Remark 3.1. Because of (2.8) the cdf (3.5) may be written as a linear combination of the fundamental solutions (2.6).

Definition 3.1. Let J be a homogeneous Markov chain on a finite state space with parameters (a, Q) satisfying (3.1)-(3.3). We call the pair $M = (a, Q)$ a *failure model*.

In the analysis of phase-type distributions it is common to talk about equivalent representations since different representations may have the same distribution of the time to absorption. Finding equivalent representations with a minimal number of states has been the topic of wide research (cf. Neuts [47]). But unlike in these attempts to simplify PH models, in our model we want to give every single state a physical description. Every state is important for the maintenance policies introduced afterwards. Only simple manipulations like permutation of states will not change the properties of the model that we are interested in.

Definition 3.2. Let $M^{(1)} = (a^{(1)}, Q^{(1)})$ and $M^{(2)} = (a^{(2)}, Q^{(2)})$ be two failure models having the same number of working states m and respective numbers of failure states $n^{(k)} - m$, $k = 1, 2$. For a given working state $i \in \{1, ..., m\}$ we define the sets of essential failure states

$$I_i^{(k)} := \{j \in \{m+1, ..., n^{(k)}\} : q_{ij}^{(k)} > 0\} \ , \ k = 1, 2 \ . \tag{3.6}$$

We call the failure models $M^{(1)}$ and $M^{(2)}$ equivalent (write $M^{(1)} \simeq M^{(2)}$) if

(i)

$$\exists \text{ bijection } g : \{1, ..., m\} \to \{1, ..., m\} :$$
$$a_i^{(1)} = a_{g(i)}^{(2)}, q_{ij}^{(1)} = q_{g(i)g(j)}^{(2)} \ \forall \ i, j = 1, ..., m.$$

(ii)

$$\forall i = 1, ..., m \ \exists \text{ bijection } h_i : I_i^{(1)} \to I_{g(i)}^{(2)} :$$
$$q_{ij}^{(1)} = q_{g(i)h_i(j)}^{(2)} \ \forall \ j \in I_i^{(1)}.$$

Remark 3.2. Since Definition 3.2 only uses bijections and equality of integers the relation $M^{(1)} \simeq M^{(2)}$ is, in fact, an equivalence relation.

We give an example to illustrate equivalent failure models:

Example 3.1. We consider the model $M^{(1)}$ with $m = 3$, $n = 6$ and

$$a^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, Q^{(1)} = \begin{pmatrix} -5 & 3 & 1 & 1 & 0 & 0 \\ 0 & -4 & 2 & 1 & 1 & 0 \\ 0 & 0 & -3 & 0 & 1 & 2 \end{pmatrix} .$$

Applying a bijection on the working states yields an equivalent failure model $M^{(2)}$ with $m = 3$, $n = 6$ and

$$a^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, Q^{(2)} = \begin{pmatrix} -4 & 2 & 0 & 1 & 1 & 0 \\ 0 & -3 & 0 & 0 & 1 & 2 \\ 3 & 1 & -5 & 1 & 0 & 0 \end{pmatrix} .$$

The minimal number of failure states equals $\max\{|I_i^{(1)}|, i = 1, ..., m\}$, cf. (3.6). For the failure model $M^{(1)}$ an equivalent failure model with the minimal number of failure states is given by $M^{(3)}$ with $m = 3$, $n = 5$ and

$$a^{(3)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, Q^{(3)} = \begin{pmatrix} -5 & 3 & 1 & 1 & 0 \\ 0 & -4 & 2 & 1 & 1 \\ 0 & 0 & -3 & 1 & 2 \end{pmatrix} .$$

The maximal number of failure states (with no superfluous states) equals $|I_1^{(1)}| + ...|I_m^{(1)}|$. For the failure model $M^{(1)}$ an equivalent failure model with the maximal number of failure states is given by $M^{(4)}$ with $m = 3$, $n = 8$ and

$$a^{(4)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, Q^{(4)} = \begin{pmatrix} -5 & 3 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -4 & 2 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 & 0 & 1 & 2 \end{pmatrix} .$$

Since in this Markov model the behavior of the system only depends on the current state, it would not make sense if the severeness of a failure depended on more than the failure state J_τ and the level of damage just before failure $J_{\tau-}$. It would be convenient to call the pair $(J_{\tau-}, J_\tau)$ the failure type, but we may reduce the severeness of a failure to J_τ only:

Remark 3.3. In a model with n states we have $m(n-m)$ different failure types. For every such a model there exists an equivalent model with at most $m(n-m)$ failure states such that each failure state has an unique predecessor. So we may deduce the state before failure from the failure state. This equivalent model is a model with the maximal number of failure states. Due to simplification of some calculations we henceforth only consider models with the maximal number of failure states, although the verbal description of some failure states might be the same. Now $J_{\tau-}$ is specified by J_τ and it is enough to call only J_τ the failure type.

In the next example we show that τ and J_τ are in general not independent.

Example 3.2. We consider the failure model (a, Q) with

$$a = e_1 \ , \ Q = \begin{pmatrix} -2 & 1 & 1 & 0 \\ 0 & -2 & 0 & 2 \end{pmatrix} .$$

Now we calculate with Example 2.6

$$e^{tQ_m} = \begin{pmatrix} e^{-2t} & te^{-2t} \\ 0 & e^{-2t} \end{pmatrix} \ \Rightarrow \ \mathrm{P}(\tau > t) = (1+t)e^{-2t} .$$

If we condition on $\{J_\tau = 3\}$ we know that τ is the first jump-time of the chain and hence exponentially distributed with rate 2.

$$\mathrm{P}(\tau > t | J_\tau = 3) = e^{-2t} \neq \mathrm{P}(\tau > t) .$$

This shows that τ and J_τ are not independent.

An important special case are models that cannot return to a previous condition without a maintenance action. We define these acyclic models.

Definition 3.3. We call a failure model M *acyclic* if there exists an equivalent model $M^* \simeq M$ whose intensity matrix Q^* is upper triangular.[1]

[1]Bobbio et al. [18] used the term *Triangular Multistate Homogeneous Markov Model* for these acyclic models. They derived an estimating procedure for the cdf (3.5) using only the observed failure times.

3.2 Failure rates

We know from Lemma 2.6 that the failure intensity of τ from (3.4) equals

$$\lambda(t) = \frac{-a' e^{tQ_m} Q_m \mathbf{1}}{a' e^{tQ_m} \mathbf{1}} \ , \quad t \geq 0 \ . \tag{3.7}$$

We want to investigate our failure model with regard to the properties of the failure intensity and recall (2.1)

$$\mathrm{P}(\tau \leq t + h | \tau > t) = h \cdot \lambda(t) + o(h) \ , \quad h \to 0 \ .$$

Conditioning on J_t yields (since $\mathrm{P}(J_t > m | \tau > t) = 0$)

$$\begin{aligned} \mathrm{P}(\tau \leq t + h | \tau > t) &= \sum_{i=1}^{n} \mathrm{P}(\tau \leq t + h | \tau > t, J_t = i) \mathrm{P}(J_t = i | \tau > t) \\ &= \sum_{i=1}^{m} \mathrm{P}(\tau \leq t + h | \tau > t, J_t = i) \mathrm{P}(J_t = i | \tau > t) \ . \end{aligned} \tag{3.8}$$

Now we may define for $i = 1, ..., m$

$$\mathrm{P}(\tau \leq t + h | \tau > t, J_t = i) =: h \cdot \tilde{\lambda}_i(t) + o(h) \ , \quad h \to 0 \ ,$$

where $\tilde{\lambda}_i(t)$ are conditional (state-dependent) intensities and, by construction

$$\tilde{\lambda}_i(t) = \sum_{k=m+1}^{n} q_{ik} \ , \text{ resp. } \tilde{\lambda}(t) = \begin{pmatrix} \tilde{\lambda}_1(t) \\ \vdots \\ \tilde{\lambda}_m(t) \end{pmatrix} = -Q_m \mathbf{1} \ .$$

Taking the limit $h \to 0$ in (3.8) yields

$$\lambda(t) = \sum_{i=1}^{m} \tilde{\lambda}_i(t) \mathrm{P}(J_t = i | \tau > t) \ , \ t \geq 0 \ . \tag{3.9}$$

The conditional failure intensities $\tilde{\lambda}(t)$ do not depend on t, so we just write $\tilde{\lambda}$. We may use these intensities to define the stochastic failure intensity process

$$\overline{\lambda}(t) := \sum_{i=1}^{m} \tilde{\lambda}_i \mathbb{1}_{\{i\}}(J_t) \ , \quad t \geq 0 \ . \tag{3.10}$$

Lemma 3.1. The unconditional failure rate $\lambda(t)$ equals the conditional expectation of the process (3.10) given $\tau > t$.

Proof:

$$\mathrm{E}\left[\overline{\lambda}(t)|\tau > t\right] = \sum_{i=1}^{m} \tilde{\lambda}_i \mathrm{E}\left[\mathbb{1}_{\{i\}}(J_t)|\tau > t\right] = \sum_{i=1}^{m} \tilde{\lambda}_i \mathrm{P}(J_t = i|\tau > t) = \lambda(t) \ .$$

□

Remark 3.4. Consider a model for the time to failure like in Example 3.1 in Finkelstein [26]. The stochastic intensity of a time to failure $\tilde{\tau}$ equals $\lambda(t, Z)$, $t \geq 0$ and given $Z = z$, the failure rate of $\tilde{\tau}$ equals $\lambda(t, z)$, $t \geq 0$. Now the exponential formula (2.2) holds

$$\mathrm{P}(\tilde{\tau} > t|Z = z) = \exp\left\{-\int_0^t \lambda(u, z) \ \mathrm{d}u\right\}$$

and by Theorem 6.2 from Finkelstein [26] the failure rate of $\tilde{\tau}$ is smaller than the mixture $\mathrm{E}(\lambda(t, Z))$. The stochastic failure intensity process $\overline{\lambda}(t)$ from (3.10) in our model behaves different. The failure time τ and $\overline{\lambda}(t)$ depend in the meaning of

$$\tau = \sup\{t \geq 0 : \overline{\lambda}(t) > 0\} \ , \ \text{P-a.s.} \ .$$

The exponential formula (2.2) does not hold for $\overline{\lambda}(t)$ since

$$\int_0^\infty \overline{\lambda}(t)dt < \infty \ , \ \text{P-a.s.} \ ,$$

and concerning the failure rate $\lambda(t)$, it follows from Lemma 3.1 that

$$\lambda(t) > \mathrm{E}(\overline{\lambda}(t)) \ \ , \ t > 0 \ , \ \lambda(0) = \mathrm{E}(\overline{\lambda}(0)) \ .$$

Remark 3.5. It follows from (3.9) that the failure rate $\lambda(t)$ is bounded

$$\min_{i=1,...,m} \tilde{\lambda}_i \leq \lambda(t) \leq \max_{i=1,...,m} \tilde{\lambda}_i \ , \ t \geq 0 \ .$$

Remark 3.6. As pointed out by Neuts [47] Section 2.3, the density function of a PH-distribution with an irreducible matrix Q_m is asymptotically exponential, which means that the failure rate $\lambda(t)$ converges to some value $\lambda^* > 0$ and with (3.9) it holds

$$\lambda^* = \lim_{t\to\infty} \sum_{i=1}^{m} \tilde{\lambda}_i \mathrm{P}(J_t = i|\tau > t) \ .$$

The conditional probabilities $\mathrm{P}(J_t = i|\tau > t)$ converge for all initial distributions a to a unique probability vector $y^* \in \mathbb{R}^m$ (the so-called quasi-stationary distribution) with

$$\lim_{t\to\infty} \mathrm{P}(J_t = i|\tau > t) = y_i^* \ , \quad i = 1, ..., m. \tag{3.11}$$

An irreducible matrix Q_m has an unique eigenvalue λ_0 with maximal real part. The asymptotic failure rate λ^* equals $-\lambda_0$ and the quasi-stationary distribution y^* is obtained as the unique probability vector that solves

$$y'Q_m = \lambda_0 y' \ , \tag{3.12}$$

hence y^* is a left-eigenvector for λ_0. Doorn and Pollett [24] investigated the reducible case. A sufficient criterion for the existence of a unique quasi-stationary distribution is, that $\mathrm{ind}(Q_m - \lambda_0 \mathbf{I}) = 1$ (λ_0 has the geometric multiplicity one).

Example 3.3. If $m = 1$ we have $\mathrm{P}(J_t = 1|\tau > t) = 1$ and $\overline{\lambda}(t) = -q_{11}$. Hence $\lambda(t) = -q_{11}$ which means that τ is exponentially distributed.

Example 3.4. If there is a constant $b > 0$ such that $\tilde{\lambda}_i = b$ for all $i = 1, ..., m$ we have

$$\overline{\lambda}(t) = \sum_{i=1}^{m} \tilde{\lambda}_i \mathrm{P}(J_t = i|\tau > t) = b \ .$$

So again, $\lambda(t) = b$ is constant and τ is exponentially distributed.

Example 3.5. We consider the case $m = 2$. We may use the matrix exponentials from Example 2.6 for to obtain the respective failure rates.

i) $a = e_1$, $Q_2 = \begin{pmatrix} q_{11} & q_{12} \\ 0 & q_{11} \end{pmatrix}$ with $-q_{11} \geq q_{12} > 0$. Now

$$\lambda(t) = -\frac{q_{12}}{1 + q_{12}t} - q_{11} \ , \quad \lambda'(t) = \frac{q_{12}^2}{(1 + q_{12}t)^2} > 0 \ , \ \forall \ t \geq 0 \ .$$

So τ is IFR, and since $\lambda(t)$ converges to $-q_{11}$ and because of (3.9) there exists the unique quasi-stationary distribution $y^* = e_2$ in the sense of (3.11).

ii) $a = e_1$, $Q_2 = \begin{pmatrix} q_{11} & q_{12} \\ 0 & q_{22} \end{pmatrix}$ with $-q_{11} \geq q_{12} > 0$, $q_{11} \neq q_{22} < 0$. Now

$$\lambda(t) = -\frac{q_{11}e^{q_{11}t}(q_{11} + q_{12} - q_{22}) - q_{12}q_{22}e^{q_{22}t}}{e^{q_{11}t}(q_{11} + q_{12} - q_{22}) - q_{12}e^{q_{22}t}} \ ,$$
$$\lambda'(t) = g(t)(q_{11} + q_{12} - q_{22}) \text{ with } g(t) > 0 \ \forall \ t \geq 0 \ .$$

So τ is $\left\{ \begin{array}{c} \text{IFR} \\ \text{DFR} \\ \sim \text{Exp} \end{array} \right\} \Leftrightarrow q_{11} + q_{12} \left\{ \begin{array}{c} > \\ < \\ = \end{array} \right\} q_{22}$. Again, $\lambda(t)$ converges:
For $q_{11} < q_{22}$ we have

$$\lim_{t\to\infty} \lambda(t) = -q_{22} \text{ and } y^* = e_2 \ .$$

For $q_{11} > q_{22}$ we have

$$\lim_{t\to\infty} \lambda(t) = -q_{11} \text{ and } y^* = \left(\frac{q_{11} - q_{22}}{q_{11} + q_{12} - q_{22}}, \frac{q_{12}}{q_{11} + q_{12} - q_{22}} \right)' \ .$$

iii) $a = e_1$, $Q_2 = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}$ with $-q_{11} \geq q_{12} > 0$, $-q_{22} > q_{21} > 0$.
Lengthy, but straightforward calculations yield

$$\lambda'(t) = g(t)(q_{11} + q_{12} - (q_{22} + q_{21})) \text{ with } g(t) > 0 \ \forall \ t \geq 0 \ .$$

So τ is $\left\{ \begin{array}{c} \text{IFR} \\ \text{DFR} \\ \sim \text{Exp} \end{array} \right\} \Leftrightarrow q_{11} + q_{12} \left\{ \begin{array}{c} > \\ < \\ = \end{array} \right\} q_{22} + q_{21}$. The matrix Q_2 is irreducible and (3.12) is solved by $u = (u_1, u_2)'$ with

$$u_1 = \frac{q_{21}}{q_{21} + \lambda_1 - q_{11}} \ , \ u_2 = \frac{\lambda_1 - q_{11}}{q_{21} + \lambda_1 - q_{11}} \ .$$

The failure rate converges, of course, to $-\lambda_1$

$$\lambda^* = \lim_{t\to\infty} \sum_{i=1}^{2} \tilde{\lambda}_i \mathrm{P}(J_t = i | \tau > t) = \sum_{i=1}^{2} \tilde{\lambda}_i u_i = -\lambda_1 \ ,$$

where λ_1 is the larger one of the eigenvalues from Example 2.6.

For $m = 2$ the sign of the failure rate $\lambda(t)$ only depends on the conditional failure intensities $\tilde{\lambda}_1, \tilde{\lambda}_2$, since τ is $\left\{ \begin{array}{c} \text{IFR} \\ \text{DFR} \end{array} \right\} \Leftrightarrow \tilde{\lambda}_1 \left\{ \begin{array}{c} \leq \\ \geq \end{array} \right\} \tilde{\lambda}_2$.
For $m > 2$ there are of course other possible shapes of $\lambda(t)$ like in the next examples.

Example 3.6. Let $m = 3$, $a = e_1$, $Q_3 = \begin{pmatrix} -10 & 5 & 0 \\ 0 & -10 & 9 \\ 0 & 0 & -10 \end{pmatrix}$. This is an acyclic model with conditional failure intensities $\tilde{\lambda} = \begin{pmatrix} 5 \\ 1 \\ 10 \end{pmatrix}$. The failure rate $\lambda(t)$ may be calculated:

$$e^{Q_3 t} = e^{-10t}\left(I + Q_3 t + 10tI + \frac{1}{2}(Q_3 t + 10tI)^2\right) = e^{-10t}\begin{pmatrix} 1 & 5t & 22.5t^2 \\ 0 & 1 & 9t \\ 0 & 0 & 1 \end{pmatrix}$$

$$\Rightarrow \lambda(t) = \frac{10(1 + t + 45t^2)}{2 + 10t + 45t^2} .$$

Now τ has a bathtub-shaped failure rate.

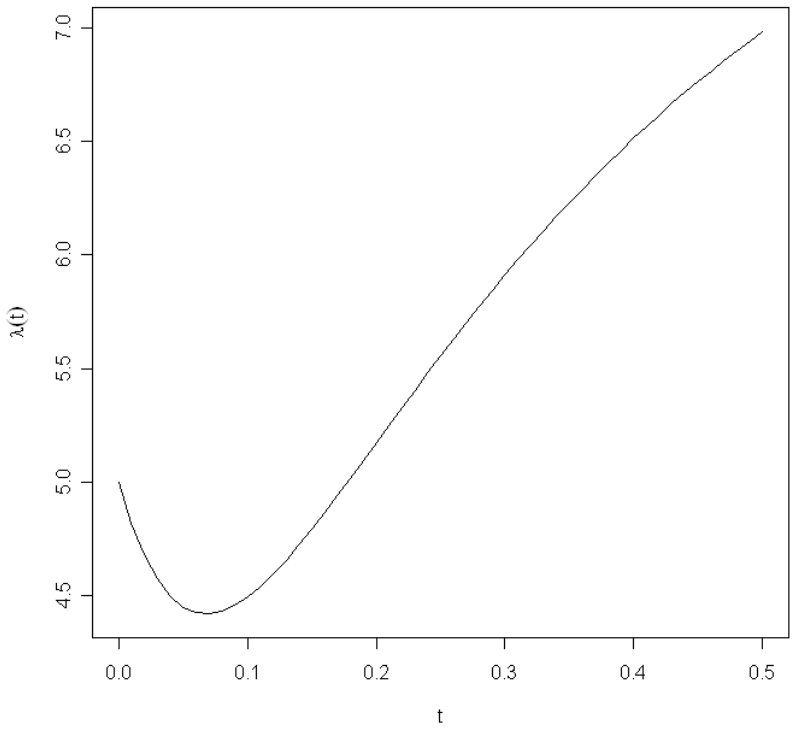


Figure 3.1: Failure rate in Example 3.6

The parameters in the above example have been chosen in order to obtain a stochastic failure rate process (3.10) that is first decreasing and later increasing, but this does not generally imply a bathtub shaped failure rate. It is not straightforward seen from the conditional failure intensities $\tilde{\lambda}$ what $\lambda(t)$ might look like. Aalen and Gjessing [3] write "The shape of the hazard rate is in many cases determined by how the starting distribution of the process relates to the quasi-stationary one, whether it is closer to or further apart from the absorbing state." A second approach to a bathtub-shaped failure rate is the burn-in phenomenon.

Example 3.7. Assume that a system usually starts well conditioned but with a small probability it starts in a very bad condition. The corresponding mixture of probability measures might be modeled by $M = (a, Q)$ with

$$a = \begin{pmatrix} p \\ 0 \\ 1-p \end{pmatrix} \ , \quad Q_3 = \begin{pmatrix} -2 & 1 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & -20 \end{pmatrix} \ , \quad p \in (0,1) \ .$$

Again we obtain with Example 2.6

$$e^{Q_3 t} = \begin{pmatrix} e^{-2t} & \frac{1}{3}e^{-2t} - \frac{1}{3}e^{-5t} & 0 \\ 0 & e^{-5t} & 0 \\ 0 & 0 & e^{-20t} \end{pmatrix} \ .$$

Hence by (3.7) we have

$$\lambda(t) = \frac{p(8e^{-2t} - 5e^{-5t}) + 60(1-p)e^{-20t}}{p(4e^{-2t} - e^{-5t}) + 3(1-p)e^{-20t}} \ , \ t \geq 0 \ .$$

For e.g. $p = 0.95$ the failure rate is bathtub-shaped (Figure 3.2).

The Examples 3.6 and 3.7 satisfy the bathtub property ii) from Definition 2.15 but they are not really convincing, since the part of the failure rate that looks similar to the bathtub curve in Figure 2.1 ($0 \leq t \leq 0.2$ in Figure 3.1, $0 \leq t \leq 0.6$ in Figure 3.2) describes only the smallest 60% of failure times in Example 3.6 (resp. 63% in Example 3.7). We obtain a more complete description of the failure times by introducing more stages.

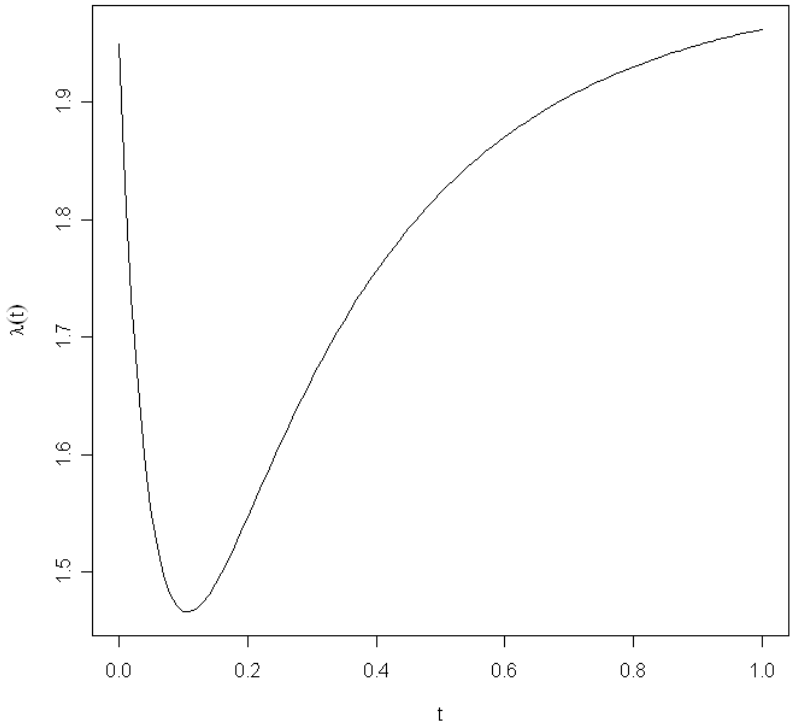


Figure 3.2: Failure rate in Example 3.7

Example 3.8. We consider a system pretty similar to that in Example 3.7 with $m = 10$ and

$$a = \begin{pmatrix} 0.95 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0.05 \end{pmatrix}, \quad Q_{10} = \begin{pmatrix} -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 \end{pmatrix}.$$

The failure rate looks like

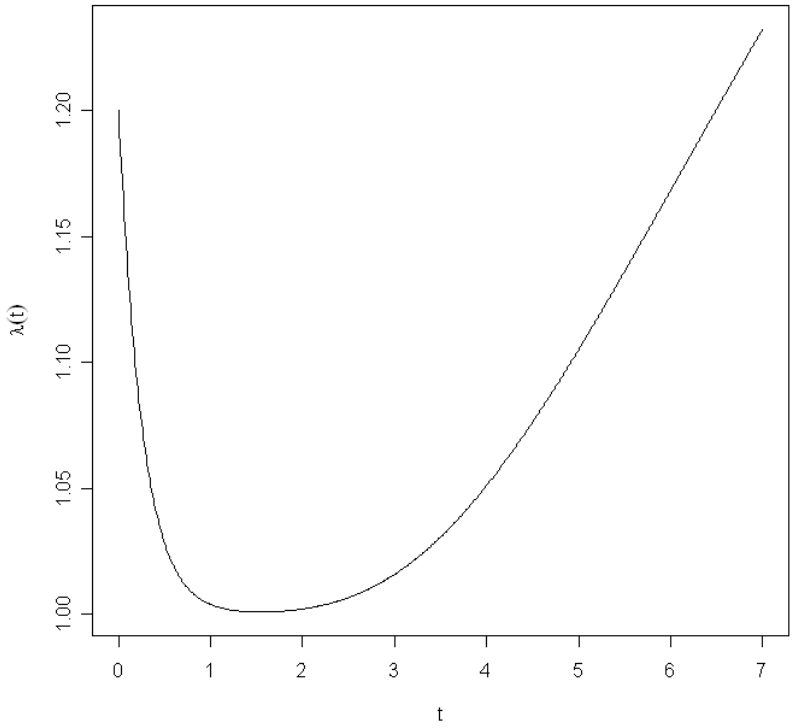


Figure 3.3: Failure rate in Example 3.8

Here the smallest $99,9\%$ of failure times are in the interval $[0,7]$ and so almost all failure times may be described by a failure rate which resembles the bathtub curve in Figure 2.1.

In the next examples we show that the monotony of the intensity process $\overline{\lambda}(t)$ does not imply the monotony of the failure rate $\lambda(t)$ and vice versa.

Example 3.9. Consider the failure model with $m = 3$ and

$$a = e_1 \ , \ Q_3 = \begin{pmatrix} -2 & 1 & 0 \\ 0 & -6 & 3 \\ 0 & 0 & -2 \end{pmatrix} \ .$$

We calculate the failure rate

$$\lambda(t) = \frac{22 + 24t - 6e^{-4t}}{17 + 12t - e^{-4t}} \ \ , \ \lambda'(t) = \frac{16(9 + 23e^{-4t} + 12te^{-4t})}{(17 + 12t - e^{-4t})^2} > 0 \ \ , \ t \geq 0 \ .$$

Although the intensity process (3.10) is not monotone with probability 0.25 (this happens when the MC visits all 3 working states before absorption), the failure rate is monotone increasing.

Example 3.10. Consider the failure model with $m = 3$ and

$$a = e_1 \ , \ Q_3 = \begin{pmatrix} -3 & 1 & 1 \\ 0 & -2 & 0 \\ 0 & 0 & -10 \end{pmatrix} \ .$$

We calculate the failure rate

$$\lambda(t) = \frac{14 + 3e^{-t} - 10e^{-8t}}{7 + e^{-t} - e^{-8t}} \ , \ \lambda'(t) = -\frac{7(e^{-t} - 64e^{-8t} - 7e^{-9t})}{(7 + e^{-t} - e^{-8t})^2} \ , \ t \geq 0 \ .$$

The intensity process is a.s. monotone increasing, but the failure rate $\lambda(t)$ is monotone decreasing for $t \geq 0.604$.

A non-trivial class of failure models where one knows that the failure rate is increasing is a class of skip-free models (cf. Aalen [2], Kijima [35]).

3.3 Information level

The failure rate process (3.10) provides accurate information about the failure behavior of a working system, but for its evaluation we have to be able to observe the MC J. In some applications the user has only partial information about the condition of the system or even no information at all. We introduce three information levels (the three respective filtrations $\mathfrak{F}^1, \mathfrak{F}^2, \mathfrak{F}^3$) which allow appealing interpretations.

Definition 3.4. In the concept of failure models on the probability space $(\Omega, \mathfrak{A}, \mathrm{P})$ we call a filtration $\mathfrak{F} = (\mathfrak{F})_{t \geq 0}$ of $\mathfrak{A}$ *information level*.

Example 3.11. No information about the state of a running system means that we only may observe the failure time τ from (3.4). In this case $\mathfrak{F}^1$ is given by the canonical filtration of the process $(\mathbb{1}_{[0,t]}(\tau))_{t \geq 0}$

$$\mathfrak{F}^1_t = \sigma\left(\mathbb{1}_{[0,s]}(\tau) : 0 \leq s \leq t\right) \ , \ t \geq 0 \ .$$

Example 3.12. Let us assume that the user may observe whether a running system is in a critical state or not (e.g. signalized by warning lights). So there is a partition $A_1, ..., A_l$ of $\{1, ..., m\}$ and we may decide in which block

A_i the current state of the system is. But if $|A_i| > 1$ we may not determine the state. $\mathfrak{F}^2$ is defined via

$$\mathfrak{F}_t^2 = \sigma\left(\mathbb{1}_{A_i}(J_s) : 0 \leq s \leq t, i = 1, ..., l\right) \ , \ t \geq 0 \ .$$

Of course the failure time is observed because

$$\sum_{i=1}^{l} \mathbb{1}_{A_i}(J_t) = \mathbb{1}_{(t,\infty)}(\tau) \ , \ t \geq 0 \ .$$

Example 3.13. All information about J is given by $\mathfrak{F}^J$. Any filtration that covers $\mathfrak{F}^J$ also contains this information but for convenience we set $\mathfrak{F}^3 = \mathfrak{F}^J$

$$\mathfrak{F}_t^3 = \sigma(J_s : 0 \leq s \leq t) = \sigma(\mathbb{1}_{\{i\}}(J_s) : 0 \leq s \leq t, \ i = 1, ..., m) \ , \ t \geq 0 \ .$$

If we compare the filtrations from Examples 3.11 - 3.13 we obtain

$$\mathfrak{F}_t^1 \subseteq \mathfrak{F}_t^2 \subseteq \mathfrak{F}_t^3 \ \ , \ t \geq 0 \ .$$

If $l = 1$ we have $\mathfrak{F}^1 = \mathfrak{F}^2$, if $l = m$ we have $\mathfrak{F}^2 = \mathfrak{F}^3$ and if $m = l = 1$ we have $\mathfrak{F}^1 = \mathfrak{F}^2 = \mathfrak{F}^3$.

3.4 SSM representation

We want to derive the SSM representations of the indicator point process Z of the failure time τ

$$Z_t = \mathbb{1}_{[0,t]}(\tau) \ \ , \ t \geq 0 \tag{3.13}$$

with respect to the filtrations $\mathfrak{F}^1$ and $\mathfrak{F}^3$ from the Examples 3.11 and 3.13. The process Z is the counting process belonging to the point process $(T_i)_{i \in \mathbb{N}}$ with $T_1 = \tau$, $T_i = \infty$, $i \geq 2$. Since $\mathfrak{F}^3$ is the bigger one of the two filtrations, we first determine the $\mathfrak{F}^3$-intensity from (2.17) of Z and then use Theorem 2.6 to obtain the $\mathfrak{F}^1$-intensity of Z.

Lemma 3.2. Let τ be the time to failure (3.4) and Z its indicator point process (3.13). Let $\mathfrak{F}^3$ be the canonical filtration with respect to the underlying MC J. Now Z has a $\mathfrak{F}^3$-SSM representation

$$Z_t = \int_0^t \lambda_s^3 \mathbb{1}_{(s,\infty)}(\tau) \mathrm{d}s + M_t \ \ , \ t \geq 0 \ ,$$

where $\lambda_s^3 = \overline{\lambda}(s)$ from (3.10) and $M = (M_t)_{t \geq 0} \in \mathcal{M}(\mathfrak{F}^3)$.

This result is a standard example for the compensator of a so-called piecewise deterministic Markov process (cf. Section 7.5 in Last and Brandt [37] or Section 7.2 in Jacobsen [31]). Nevertheless, we present a proof and use only the definition of the conditional expectation (Definition 2.20).

Proof: We have to show that $M = (M_t)_{t\geq 0} \in \mathcal{M}(\mathfrak{F}^3)$. Obviously, M is cadlag and $M_0 = 0$. Now for $0 \leq s < t$ holds

$$\begin{aligned}\mathrm{E}(M_t|\mathfrak{F}_s^3) &= \mathrm{E}\left(\mathbb{1}_{[0,t]}(\tau)|\mathfrak{F}_s^3\right) - \mathrm{E}\left(\int_0^t \lambda_u^3 \mathbb{1}_{(u,\infty)}(\tau)\mathrm{d}u \mid \mathfrak{F}_s^3\right) \\ &= M_s + \mathrm{E}\left(\mathbb{1}_{(s,t]}(\tau)|\mathfrak{F}_s^3\right) - \mathrm{E}\left(\int_s^t \lambda_u^3 \mathbb{1}_{(u,\infty)}(\tau)\mathrm{d}u \mid \mathfrak{F}_s^3\right) \ .\end{aligned}$$

The σ-algebra $\mathfrak{F}_s^3$ is generated by the events $\bigcap_{i=1}^N \{J_{t_i} = k_i\}$, with arbitrary $N \in \mathbb{N}$, $0 \leq t_1 \leq ... \leq t_N \leq s$, $k_1, ..., k_N \in \{1, ..., n\}$. For any such event $A = \bigcap_{i=1}^N \{J_{t_i} = k_i\}$ with $\mathrm{P}(A) > 0$ we show that

$$\int_A \mathbb{1}_{(s,t]}(\tau)\mathrm{dP} = \int_A \int_s^t \lambda_u^3 \mathbb{1}_{(u,\infty)}(\tau)\mathrm{d}u\mathrm{dP} \ .$$

The first integral simplifies because of the Markov property (2.9)

$$\mathrm{P}(s < \tau \leq t, A) = \mathrm{P}(s < \tau \leq t|A)\mathrm{P}(A) = \mathrm{P}(s < \tau \leq t|J_{t_N} = k_N)\mathrm{P}(A) \ .$$

In the second integral we use again the Markov property (2.9) and the homogeneity of the MC

$$\begin{aligned}&\int_\Omega \int_s^t \sum_{i=1}^m \tilde{\lambda}_i \mathbb{1}_A \mathbb{1}_{\{i\}}(J_u)\mathrm{d}u\mathrm{dP} = \int_s^t \sum_{i=1}^m \tilde{\lambda}_i \mathrm{P}(J_u = i, A)\mathrm{d}u \\ &= \int_s^t \sum_{i=1}^m \tilde{\lambda}_i \mathrm{P}(J_u = i|\tau > u, A)\mathrm{P}(\tau > u|A)\mathrm{P}(A)\mathrm{d}u \\ &= \mathrm{P}(A)\int_s^t \sum_{i=1}^m \tilde{\lambda}_i \mathrm{P}(J_u = i|\tau > u, J_{t_N} = k_N)\mathrm{P}(\tau > u|J_{t_N} = k_N)\mathrm{d}u \\ &= \mathrm{P}(A)\int_s^t \lambda^{J_0 = k_N}(u - t_N)\mathrm{P}(\tau > u - t_N|J_0 = k_N)\mathrm{d}u \\ &= \mathrm{P}(A)\int_s^t f_\tau^{J_0 = k_N}(u - t_N)\mathrm{d}u \\ &= \mathrm{P}(A)\mathrm{P}(s - t_N < \tau \leq t - t_N|J_0 = k_N) = \mathrm{P}(A)\mathrm{P}(s < \tau \leq t|J_{t_N} = k_N) \ .\end{aligned}$$

This proves that both conditional expectations coincide and we obtain

$$\mathrm{E}(M_t|\mathfrak{F}_s^3) = M_s \ .$$

□

Remark 3.7. We may verify a specific version Y of $\mathrm{E}\left(\mathbb{1}_{(s,t]}(\tau)|\mathfrak{F}_s^3\right)$, namely

$$Y = \sum_{i=1}^{m} \mathrm{P}(s < \tau \leq t | J_s = i) \mathbb{1}_{\{i\}}(J_s) \ .$$

Clearly, Y is $\mathfrak{F}_s^3$-measurable. For $A \in \mathfrak{F}_s^3$ we have with Chapman-Kolmogorov (2.11)

$$\begin{aligned}
\int_A \sum_{i=1}^{m} \mathrm{P}(s < \tau \leq t | J_s = i) \mathbb{1}_{\{i\}}(J_s) \mathrm{dP} &= \sum_{i=1}^{m} \mathrm{P}(s < \tau \leq t | J_s = i) \mathrm{P}(A, J_s = i) \\
&= \sum_{i=1}^{m} \mathrm{P}(s < \tau \leq t | J_s = i) \mathrm{P}(J_s = i | A) \mathrm{P}(A) \\
&= \mathrm{P}(s < \tau \leq t | A) \mathrm{P}(A) \ .
\end{aligned}$$

Lemma 3.3. Let τ be the time to failure (3.4) and Z its indicator point process (3.13). Let $\mathfrak{F}^1$ be the canonical filtration $\mathfrak{F}^Z$ from Example 3.12. Now the $\mathfrak{F}^1$-intensity $\lambda^1 = (\lambda_t^1)_{t \geq 0}$ of Z is given by

$$\lambda_t^1 = \lambda(t) \mathbb{1}_{(t,\infty)}(\tau) \quad , \ t \geq 0 \ .$$

with $\lambda(t)$ from (3.7).

Proof: : It follows from Theorem 2.6 that

$$\lambda_t^1 = \mathrm{E}(\lambda_t^3 | \mathfrak{F}_t^1) \ .$$

Since $\lambda_t^3 = \overline{\lambda}(t) \mathbb{1}_{(t,\infty)}(\tau)$ we obtain with (2.16)

$$\lambda_t^1 = h(\{\mathbb{1}_{[0,s]}(\tau) : 0 \leq s \leq t\}) \ ,$$

and since $\lambda_t^3 = 0$ on $\{\tau \leq t\}$ this simplifies to $\lambda_t^1 = h(\mathbb{1}_{[0,t]}(\tau))$. With Lemma 3.1 we obtain

$$\lambda_t^1 = \begin{cases} \mathrm{E}(0 | \tau \leq t) & , \ \tau \leq t \\ \mathrm{E}(\overline{\lambda}(t) | \tau > t) & , \ \tau > t \end{cases} = \begin{cases} 0 & , \ \tau \leq t \\ \lambda(t) & , \ \tau > t \end{cases} \ .$$

□

Chapter 4

Maintenance actions in case of a failure

When the system has a failure, we want to repair it. Repairing means that we set the system again into a working state. The repair action depends on the failure type, since for different failure types different repair actions might be applicable. We still assume that we always have a failure model $M = (a, Q)$ with the maximal number of failure states.

4.1 Repair matrices

We assume that for every failure type we have a finite number of possible repair actions (but at least one). A repair action is described by a probability vector $r \in \mathbb{R}^m$ where r_i gives the probability that after repair the system restarts in state $i,\ i = 1, ..., m$.

Definition 4.1. Let M be a failure model with n states and m working states and for $i = 1, ..., n-m$ let $\mathcal{R}_i = \{r^{i,1}, ..., r^{i,n_i}\} \subset \mathbb{R}^m$, $n_i \in \mathbb{N}$, be a non-empty finite set of probability vectors. An matrix $R \in \mathbb{R}^{(n-m)\times m}$ whose rows $(r^{1,j_1})', ..., (r^{n-m,j_{n-m}})'$ are probability vectors $r^i \in \mathcal{R}_i$ is called *repair matrix*. We define the set

$$\mathcal{R}(M) := \{R : R \text{ is a repair matrix for model } M\}$$

of possible repair matrices. We call the pair $(M, \mathcal{R}(M))$ a *repair model*.

A repair matrix is build by choosing an appropriate repair action for every failure type. Repair is performed by one multinomial trial with respect to the probability vector belonging to the given failure type. The system

restarts in the state that results from this trial. For computational purposes we may order the repair matrices. We may identify a repair matrix $R = (r^{1,j_1}|...|r^{n-m,j_{n-m}})' \in \mathcal{R}(M)$ by its index vector $j = (j_1, ..., j_{n-m})'$. We order all possible repair matrices by the lexicographical order of the respective index vectors and write $\mathcal{R}(M) = \{R_1, ..., R_{|\mathcal{R}(M)|}\}$.

Example 4.1. Consider again the failure model $M^{(4)}$ from Example 3.1 with $n - m = 5$ failure states. Assume that we have the repair opportunities

$$\mathcal{R}_1 = \{e_1\} \ , \ \mathcal{R}_2 = \left\{ e_2, \begin{pmatrix} 0.8 \\ 0.2 \\ 0 \end{pmatrix} \right\} \ , \ \mathcal{R}_3 = \left\{ e_2, \begin{pmatrix} 0.6 \\ 0.3 \\ 0.1 \end{pmatrix} \right\} \ ,$$

$$\mathcal{R}_4 = \left\{ e_3, \begin{pmatrix} 0.5 \\ 0.3 \\ 0.2 \end{pmatrix} \right\} \ , \ \mathcal{R}_5 = \left\{ e_3, \begin{pmatrix} 0.2 \\ 0.6 \\ 0.2 \end{pmatrix} \right\} \ .$$

Since $n_1 = 1, \ n_2, n_3, n_4, n_5 = 2$ we may construct $|\mathcal{R}(M)| = 16$ different repair matrices. One choice could be

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0.8 & 0.2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.2 & 0.6 & 0.2 \end{pmatrix} \ .$$

The corresponding index vector equals $j = (1, 2, 1, 1, 2)$ and the lexicographical order of $\mathcal{R}(M)$ is given by

1	1	1	1	1	1	9	1	2	1	1	1
2	1	1	1	1	2	10	1	2	1	1	2
3	1	1	1	2	1	11	1	2	1	2	1
4	1	1	1	2	2	12	1	2	1	2	2
5	1	1	2	1	1	13	1	2	2	1	1
6	1	1	2	1	2	14	1	2	2	1	2
7	1	1	2	2	1	15	1	2	2	2	1
8	1	1	2	2	2	16	1	2	2	2	2

Hence the index vector $j = (1, 2, 1, 1, 2)$ belongs to the repair matrix R_{10}.

Definition 4.2. We define a marked point process $(\tau^R, U^R) = (\tau_k^R, U_k^R)_{k\in\mathbb{N}}$ and a process $S^R = (S_t^R)_{t\geq 0}$ using a repair matrix $R = (r^1|...|r^{n-m})' \in \mathcal{R}(M)$ and copies $J^{(k)}$ of the MC J from Definition 3.1 with different initial distributions $a^{(k)} \in \mathbb{R}^m$, $k \in \mathbb{N}$ (let $\tau_0^R = 0$):

- $a^{(1)} := a$, $\tau_1^R := \inf\{t \geq 0 : J_t^{(1)} > m\}$, $U_1^R := J_{\tau_1^R}^{(1)}$,
- for $k \geq 2$:

$$\begin{aligned} S_t^R &:= J_{t-\tau_{k-2}^R}^{(k-1)} , && \text{for } \tau_{k-2}^R \leq t < \tau_{k-1}^R , \\ a^{(k)} &:= r^{U_{k-1}^R - m} , \\ \tau_k^R &:= \tau_{k-1}^R + \inf\{t \geq 0 : J_t^{(k)} > m\} , \\ U_k^R &:= J_{\tau_k^R - \tau_{k-1}^R}^{(k)} . \end{aligned}$$

Now (τ^R, U^R) gives the failure times and the corresponding failure types, whereas S^R gives the actual state of the repaired system.

Remark 4.1. The construction in Definition 4.2 generalizes the construction of a PH-renewal process (cf. Neuts [46] and [47] Section 2.4).

Remark 4.2. The process $(U_k^R)_{k\in\mathbb{N}}$ is a discrete-time MC with index set $\mathbb{N}$ and state space $\{m+1, ..., n\}$. The transition probabilities are given by

$$\mathrm{P}(U_{k+1}^R = j | U_k^R = i) = (r^{i-m, j_{i-m}})' B e_{j-m} \quad i, j = m+1, ..., n ,$$

where $(r^{i-m,j_{i-m}})'$ is the $(i-m)$-th row of R and B is the matrix of absorption probabilities from Lemma 2.2. If $U_k = i$, the inter-failure time $\tau_{k+1}^R - \tau_k^R$ is PH-distributed with representation $(r^{i-m,j_{i-m}}, Q_m)$, $i = m+1, ..., n$, $k \in \mathbb{N}$. Hence the process $\tilde{U^R} = (\tilde{U^R}_t)_{t\geq\tau_1^R}$ with

$$\tilde{U^R}_t = \sum_{k=1}^{\infty} U_k^R \mathbb{1}_{[\tau_k^R, \tau_{k+1}^R)}(t) \quad , \ t \geq \tau_1^R$$

is a semi-Markov process (cf. Limnios and Oprisan [38]).

Theorem 4.1. The process $S^R = (S_t^R)_{t\geq 0}$ from Definition 4.2 is a homogeneous MC with state space $\{1, ..., m\}$, initial distribution a and intensity matrix $\Psi^R = \left(\psi_{ij}^R\right)_{i,j=1,...,m} \in \mathbb{R}^{m\times m}$ with

$$\Psi^R = Q_m + Q_0 \cdot R .$$

The distribution of S_t^R is hence given by:

$$\mathrm{P}(S_t^R = i) = a' e^{t\Psi^R} e_i \ , \ t \geq 0 .$$

Proof: Obviously, a is the initial distribution of S^R. First we consider the case $\psi_{ii} = 0$. This occurs if

$$-q_{ii} = \sum_{l=m+1}^{n} q_{il} R_{l-m,i} = \sum_{l=m+1}^{n} q_{il} \ ,$$

which means that if the MC J jumps from state i into a failure state, the system always restarts in state i. Also since $e_i' Q \mathbf{1} = 0$ the MC J may not jump from state i into another working state $j \neq i$. Applying the repair matrix R means that i is now an absorbing state.
Consider now a fixed (non-absorbing) starting point $S_0^R = i$ with $\psi_{ii}^R \neq 0$. We have to show that for the first jump-time $Y = \inf\{t > 0 : S_t^R \neq i\}$ holds

i) $\mathrm{P}^{Y|S_0^R=i} = \mathrm{Exp}(-\psi_{ii}^R)$,

ii) $\mathrm{P}(S_Y^R = j | S_0^R = i) = -\dfrac{\psi_{ij}^R}{\psi_{ii}^R}$ for $i \neq j$.

to i) Let N be the number of failures τ_k^R until Y ($\tau_N^R \leq Y < \tau_{N+1}^R$). Now $N + 1 \sim \mathrm{Geo}(p)$, where

$$1 - p = \sum_{l=m+1}^{n} \frac{q_{il}}{-q_{ii}} R_{l-m,i}$$

is the probability of having a failure with S^R remaining in state i. The distribution of Y is the $(N+1)$-th convolution of $\mathrm{Exp}(-q_{ii})$

$$\mathrm{P}^{Y|S_0^R=i} = (\mathrm{Exp}(-q_{ii}))^{*(N+1)} \ .$$

Since N is independent of the waiting times, conditioning on N yields

$$\mathrm{P}^{Y|S_0^R=i} = \sum_{k=1}^{\infty} (\mathrm{Exp}(-q_{ii}))^{*k} \mathrm{P}(N+1 = k) = \sum_{k=1}^{\infty} (\mathrm{Exp}(-q_{ii}))^{*k} p(1-p)^{k-1} \ .$$

The characteristic function of Y is

$$\begin{aligned}
\varphi_Y(t) = \mathrm{E}e^{itY} &= \sum_{k=1}^{\infty} \left(\frac{-q_{ii}}{-q_{ii} - it} \right)^k p(1-p)^{k-1} \\
&= p \cdot \frac{-q_{ii}}{-q_{ii} - it} \cdot \sum_{k=0}^{\infty} \left(\frac{-q_{ii}}{-q_{ii} - it} \right)^k (1-p)^k \\
&= p \cdot \frac{-q_{ii}}{-q_{ii} - it} \cdot \frac{1}{1 - \frac{-(1-p)q_{ii}}{-q_{ii} - it}} \\
&= p \cdot \frac{-q_{ii}}{-q_{ii} - it} \cdot \frac{-q_{ii} - it}{-pq_{ii} - it} = \frac{-pq_{ii}}{-pq_{ii} - it} \ .
\end{aligned}$$

This shows that Y is exponentially distributed with rate $-pq_{ii} = -\psi_{ii}^R$.

to ii) Obviously the rate of entering state j is given by ψ_{ij}^R and independent of Y. Since the exit rate for state i is $-\psi_{ii}^R$, the one-step transition probability is $\mathrm{P}(S_Y^R = j | S_0^R = i) = -\dfrac{\psi_{ij}^R}{\psi_{ii}^R}$ for $i \neq j$.

□

Remark 4.3. We may generalize the construction considered in Definition 4.2 and Theorem 4.1. Let $Q = \begin{pmatrix} Q_m & Q_0 \\ * & \end{pmatrix}$ be the intensity matrix of MCs $J^{(1)}, J^{(2)}, ...$ on $\{1, ..., n\}$ where $Q_m \in \mathbb{R}^{m \times m}$ with $1 \leq m < n$. Let

$$\delta_k := \inf\{t \geq 0 : J_t^{(k)} > m\}$$

and assume that δ_k is finite P-a.s., $k \in \mathbb{N}$. Furthermore let $R \in \mathbb{R}^{(n-m)\times m}$ be a stochastic matrix. Let the initial distribution $a^{(1)}$ of $J^{(1)}$ be concentrated on $\{1, ..., m\}$ and the respective initial distributions $a^{(k+1)}$ of $J^{(k+1)}$ be the row $J_{\delta_k}^{(k)} - m$ of R, $k \in \mathbb{N}$. Then connecting the paths of $J^{(1)}, J^{(2)}, ...$ up to the respective times $\tau_1, \tau_2, ...$ yields the process $\tilde{J}$ (let $\tau_i := \delta_0 + ... + \delta_i$, $i \in \mathbb{N}$ and $\tau_0 := 0$):

$$\tilde{J}_t := \sum_{i=0}^{\infty} J_{t-\tau_i}^{(i+1)} \mathbb{1}_{[\tau_i, \tau_{i+1})}(t) \quad , \ t \geq 0 \ .$$

Now $\tilde{J}$ is a MC with intensity matrix $Q_m + Q_0 \cdot R$ and initial distribution $a^{(1)}|_{\{1,...,m\}}$.

Example 4.2. We call a repair matrix R *determined repair matrix* if its row vectors are unit vectors

$$R = (e_{j_1}, ..., e_{j_{n-m}})' \ , \ j_1, ..., j_{n-m} \in \{1, ..., m\} \ .$$

If we use a determined repair matrix, repair is always performed with certainty. Special cases of determined repair matrices are the minimal repair matrix R for which holds

$$q_{i,j+m} > 0 \ \Rightarrow \ R_{ji} = 1 \ , \ i = 1, ..., m \ , \ j = 1, ...n - m \ ,$$

and the replacement repair matrices $R^1, ..., R^m$ with

$$R_{ji}^i = 1 \ , \ i = 1, ..., m \ , \ j = 1, ..., n - m \ . \tag{4.1}$$

A repair matrix that is not determined is called *imperfect repair matrix*. A replacement repair matrix causes the system to be repaired into the same state for each failure type. The point process of failure times is then a (delayed) renewal process.

4.2 Information level

So far we had no further restrictions on the repair matrix, but we may easily think of situations where not all repair policies are applicable. Clearly, if we are not able to observe the type of failure instantaneously we may not perform a repair policy which depends on the type of failure. This motivates us to introduce information levels for failure models.

Definition 4.3. In the concept of repair models $(M, \mathcal{R}(M))$ on the probability space $(\Omega, \mathfrak{A}, \mathrm{P})$ we call a filtration $\mathfrak{F} = (\mathfrak{F}_t)_{t\geq 0}$ of $\mathfrak{A}$ *information level*.

Now we define what we understand by applicable repair matrices.

Definition 4.4. Let $(M, \mathcal{R}(M))$ be a repair model with information level $\mathfrak{F}$ and l be maximal such that there exists a partition $A_1, ..., A_l$ of $\{m+1, ..., n\}$ with

$$\mathfrak{F}_t \supseteq \sigma\left(\mathbb{1}_{[0,s]\times A_i}(\tau_k^R, U_k^R) : 0 \leq s \leq t, i = 1, ..., l,\ k \in \mathbb{N}\ ,\ R \in \mathcal{R}(M)\right)\ ,\ t \geq 0\,.$$

A repair matrix R is *applicable* if its rows $(r^1)', ..., (r^{n-m})'$ fulfill

$$i, j \in A_k\ \Rightarrow\ r^{i-m} = r^{j-m}\ ,\ k = 1, ..., l\ .$$

So the rows of the repair matrix are the same for indistinguishable failure types.

Remark 4.4. The partition in Definition 4.4 is unique up to permutations of the sets $A_1, ..., A_l$.

We give some examples for obvious choices of information levels if we consider the process $(\tau_k^R, U_k^R)_{k\in\mathbb{N}}$ from Definition 4.2.

Example 4.3. The minimal information level we should investigate is that we only may observe the failure times $(\tau_k^R)_{k\in\mathbb{N}}$. If we were not able to observe those failure times, we could not perform any repair policy. In this case $\mathfrak{F}^1$ is given by the canonical filtration of the processes $(\mathbb{1}_{[0,t]}(\tau_k^R))_{t\geq 0}$, $k \in \mathbb{N}$ and for $t \geq 0$ we have

$$\mathfrak{F}_t^1 = \sigma\left(\mathbb{1}_{[0,s]}(\tau_k^R) : 0 \leq s \leq t,\ k \in \mathbb{N}\ ,\ R \in \mathcal{R}(M)\right)\ .$$

The applicable repair matrices are all possible repair matrices whose rows are equal. If we consider the determined repair model from Example 4.2 the applicable repair matrices are the replacement matrices $R^1, ..., R^m$ from (4.1).

Example 4.4. Since repair actions only depend on the failure type, we may perform any possible repair policy if the failure type is observed instantaneously. In this case $\mathfrak{F}^2$ is the canonical filtration of the processes $(\mathbb{1}_{[0,t]\times\{j\}}(\tau_k^R, U_k^R))_{t\geq 0}$, $j = m+1, ..., n$, $k \in \mathbb{N}$ and for $t \geq 0$ we have [1]

$$\mathfrak{F}_t^2 = \sigma\left(\mathbb{1}_{[0,s]\times\{j\}}(\tau_k^R, U_k^R) : 0 \leq s \leq t, j = m+1, ..., n,\ k \in \mathbb{N}\ ,\ R \in \mathcal{R}(M)\right)\ .$$

Example 4.5. Let $A_1, ..., A_l$ be a partition of $\{m+1, ..., n\}$. An information level $\mathfrak{F}^3$ is defined by

$$\mathfrak{F}_t^3 = \sigma\left(\mathbb{1}_{[0,s]\times A_i}(\tau_k^R, U_k^R) : 0 \leq s \leq t, i = 1, ..., l,\ k \in \mathbb{N}\ ,\ R \in \mathcal{R}(M)\right)\ , t \geq 0.$$

The applicable repair matrices are the possible repair matrices whose rows are equal on the respective sets $A_1, ..., A_l$. This information level might make sense, if one originally does not have a failure model with the maximal number of failure states. Assume that there is a failure state with two predecessors and we may only observe the failure state. If we consider an equivalent failure with the maximal number of failure states, there are now two indistinguishable failure states. We only may observe, that either of the two failures occurs.

Example 4.5 covers Example 4.3 with $l = 1$ and Example 4.4 with $l = n - m$.

4.3 Failure counting processes

If we want to quantify the performance of a repair model $(M, \mathcal{R}(M))$, the most interesting aspect is the number of failures for a given repair matrix $R \in \mathcal{R}(M)$. We define the canonical counting processes for the marked point process (τ^R, U^R) from Definition 4.2.

Definition 4.5. Let $(M, \mathcal{R}(M))$ be a repair model from Definition 4.1, $R \in \mathcal{R}(M)$ and $A \subseteq \{m+1, ..., n\}$. The process $N(R, A)$ with

$$N_t(R, A) := \sum_{k=1}^{\infty} \mathbb{1}_{[0,t]\times A}(\tau_k^R, U_k^R)\ \ ,\ t \geq 0$$

is the *canonical counting process* for the repair matrix R and the set of marks A.

[1] or any filtration that covers this filtration

We write short: $N(R,i) := N(R,\{i\})$, $i = m+1,...,n$ and $N(R) := N(R,\{m+1,...,n\})$. There is a one-to-one relation between the marked point process (τ^R, U^R) and the canonical counting processes $N(R,A)$, since

$$\tau_k^R = \inf\{t \geq 0 : N_t(R) = k\} \ , \ U_k^R = \sum_{i=m+1}^{n} i(N_{\tau_k}(R,i) - N_{\tau_k -}(R,i)) \ , \ k \in \mathbb{N} \, .$$

So, the filtrations $\mathfrak{F}^1, \mathfrak{F}^2, \mathfrak{F}^3$ from Examples 4.3 - 4.5 may be written as

$$\begin{aligned}
\mathfrak{F}_t^1 &= \sigma(N_s(R), 0 \leq s \leq t) \ , & t \geq 0 \ , \\
\mathfrak{F}_t^2 &= \sigma(N_s(R,i), 0 \leq s \leq t, i = m+1,...,n) \ , & t \geq 0 \ , \\
\mathfrak{F}_t^3 &= \sigma(N_s(R,A_i), 0 \leq s \leq t, i = 1,...,l) \ , & t \geq 0 \ .
\end{aligned}$$

4.3.1 SSM representation

We now derive the $\mathfrak{F}$-intensity of the counting processes, where

$$\mathfrak{F}_t = \sigma(N_u(R,i), i = m+1,...,n, S_u^R, 0 \leq u \leq t)$$

is the canonical filtration of the processes $N(R,i)$, $i = m+1,...,n$ and S^R from Definition 4.2.

Lemma 4.1. Let $(M, \mathcal{R}(M))$ be a repair model (Definition 4.1), $R \in \mathcal{R}(M)$ and $\mathfrak{F}_t = \sigma(N_u(R,i), i = m+1,...,n, S_u^R, 0 \leq u \leq t)$, $t \geq 0$. The $\mathfrak{F}$-intensity of the counting process $N(R,i)$ from Definition 4.5 is given by

$$\overline{\lambda^i}_t = \sum_{j=1}^{m} q_{ji} \mathbb{1}_{\{j\}}(S_t^R) \quad , \ t \geq 0 \ , \ i = m+1,...,n \ .$$

Again, this result is a standard example for the compensator of a so-called piecewise deterministic Markov process (cf. Section 7.5 in Last and Brandt [37] or Section 7.2 in Jacobsen [31]). We may see this directly from (2.18). Let $\tau^i := \inf\{u > t : N_u(R,i) = N_t(R,i) + 1\}$ be the time of the next failure of type i. Then

$$\overline{\lambda^i}_t = \lim_{h \to 0+} h^{-1} \sum_{j=1}^{m} \mathrm{P}(\tau^i \leq t + h | S_t^R = j) \mathbb{1}_{\{j\}}(S_t^R) \quad , \ t \geq 0 \ .$$

We have the $\mathfrak{F}$-SSM representation

$$N_t(R,i) = \int_0^t \overline{\lambda^i}_u \mathrm{d}u + M_t \ . \tag{4.2}$$

We may now determine the mean value function of $N(R,i)$ just by taking expectations.

Corollary 4.1. We use (4.2) for calculating the mean value functions

$$\mu(t,R,i) := \mathrm{E}(N_t(R,i)) = \int_0^t a' e^{u\Psi^R} Q_0 e_{i-m} \mathrm{d}u \ , \ t \geq 0 \ ,$$
$$\mu(t,R) := \mathrm{E}(N_t(R)) = \int_0^t a' e^{u\Psi^R} Q_0 \mathbf{1} \mathrm{d}u \ , \ t \geq 0 \ .$$

4.3.2 The distribution of the number of failures

The counting process $N(R)$ gives the total number of failures

$$N(R) = \sum_{i=m+1}^{n} N(R,i).$$

We are able to derive the distribution of $N_t(R)$. We have

$$\mathrm{P}(N_t(R)=k) = \begin{cases} \mathrm{P}(\tau_{k+1}^R > t) - \mathrm{P}(\tau_k^R > t) & , \ k \geq 1 \ , \\ \mathrm{P}(\tau_1^R > t) & , \ k = 0 \ . \end{cases} \tag{4.3}$$

Now $\mathrm{P}(\tau_1^R > t) = a' e^{tQ_m}\mathbf{1}$ and τ_2^R is the sum of the two phase-type distributed random variables τ_1^R and $\tau_2^R - \tau_1^R$ which are in general not independent, because the initial distribution of $\tau_2^R - \tau_1^R$ might depend on τ_1^R. So we may not apply Lemma 2.7 directly but the construction in Definition 4.2 yields a phase-type representation for τ_2^R.

Lemma 4.2. The random variable τ_2 from Definition 4.2 is phase-type distributed with representation

$$a^{(2)} = \begin{pmatrix} a \\ 0 \end{pmatrix} \ , \ Q^{(2)} = \begin{pmatrix} Q_m & Q_0 R \\ 0 & Q_m \end{pmatrix} \ .$$

Proof: We define the MC $\tilde{J} = \left(\tilde{J}_t\right)_{t\geq 0}$ by

$$\tilde{J}_t = \begin{cases} J_t^{(1)} & , \ 0 \leq t < \tau_1^R \\ J_{t-\tau_1^R}^{(2)} + m & , \ t \geq \tau_1^R \end{cases} .$$

The MC $\tilde{J}$ takes values in $\{1,...,n+m\}$ where $\{1,...,2m\}$ are transient states and $\{2m+1,...,n+m\}$ are absorbing states. The initial distribution $a^{(2)}$ and the intensity matrix $Q^{(2)}$ are as given above and τ_2^R is the time to absorption of $\tilde{J}$.

$$\tau_2^R = \inf\{t \geq 0 : \tilde{J}_t > 2m\} \ \Rightarrow \ \tau_2^R \sim \mathrm{PH}(a^{(2)}, Q^{(2)}) \ .$$

□

Corollary 4.2. Per induction we obtain $\tau_k^R \sim \mathrm{PH}(a^{(k)}, Q^{(k)})$ for $k \geq 3$ with

$$a^{(k)} = \begin{pmatrix} a^{(k-1)} \\ 0 \end{pmatrix} \ , \ Q^{(k)} = \left(\begin{array}{c|c} Q^{(k-1)} & \begin{matrix} 0 \\ Q_0 R \end{matrix} \\ \hline \begin{matrix} 0 & \cdots & 0 \end{matrix} & Q_m \end{array}\right) \ .$$

The explicit form is

$$a^{(k)} = \begin{pmatrix} a \\ 0 \end{pmatrix} \in \mathbb{R}^{km} \ , \ Q^{(k)} = \begin{pmatrix} Q_m & Q_0 R & 0 & \cdots & 0 \\ 0 & Q_m & Q_0 R & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & Q_0 R \\ 0 & \cdots & \cdots & 0 & Q_m \end{pmatrix} \in \mathbb{R}^{km \times km} \ .$$

Formula (4.3) now becomes

$$\mathrm{P}(N_t(R) = k) = (a^{(k+1)})' e^{tQ^{(k+1)}} \mathbf{1} - (a^{(k)})' e^{tQ^{(k)}} \mathbf{1} \ . \tag{4.4}$$

Again we have to calculate matrix exponentials. This time the dimension of the matrix is not bounded for $k \to \infty$. In the following we derive an algorithm that approximates the distribution of $N_t(R)$.

Lemma 4.3. For $k \in \mathbb{N}$ let $Q^{(k)}$ be defined as in Corollary 4.2. For $l \in \mathbb{N}$ the l-th power of $Q^{(k)}$ is given by

$$\left(Q^{(k)}\right)^l = \begin{pmatrix} A_{0,l} & A_{1,l} & \cdots & A_{k-1,l} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & A_{1,l} \\ 0 & \cdots & 0 & A_{0,l} \end{pmatrix} \ , \tag{4.5}$$

where for $j = 0, ..., m-1$

$$A_{j,l} = \sum_{x \in \{0,1\}^l, |x| = j} \prod_{i=1}^{l} Q_m^{1-x_i} (Q_0 R)^{x_i} \ .^2 \tag{4.6}$$

Proof: Let $k \in \mathbb{N}$ be fixed. The proof goes via induction on $l \in \mathbb{N}$. First we note that for $l, j \in \mathbb{N}$ the matrices $A_{j,l}$ satisfy $A_{j,l} = 0$ if $j > l$, because

[2]The matrices $A_{j,l}$ appear in the binomial formula: $(Q_m + Q_0 R)^l = A_{0,l} + ... + A_{l,l}$

j gives the number of factors Q_0R in a product with l factors. Also the recursion

$$A_{j,l+1} = A_{j,l}Q_m + A_{j-1,l}Q_0R \tag{4.7}$$

holds, because the first summand corresponds to all summands in (4.6) with $x_{l+1} = 0$ whereas the second summand corresponds to all summands with $x_{l+1} = 1$. Now the induction basis for $l = 1$ yields

$$A_{0,1} = Q_m \ , \quad A_{1,1} = Q_0R \ , \quad A_{j,1} = 0 \ , \ j \geq 2 \ .$$

For the induction step assume that (4.5) holds for some $l \in \mathbb{N}$. Now

$$\left(Q^{(k)}\right)^{l+1} = \left(Q^{(k)}\right)^{l} Q^{(k)} \ .$$

Because of the block structure $Q^{(k)} = (B_{uv})_{u,v=1,\ldots,k}$, $\left(Q^{(k)}\right)^{l} = (C_{uv})_{u,v=1,\ldots,k}$ with

$$B_{uv} = \begin{cases} Q_m & , \ u = v \\ Q_0R & , \ u+1 = v \\ 0 & , \ \text{else} \end{cases} \ , \quad C_{uv} = \begin{cases} A_{v-u,l} & , \ u \leq v \\ 0 & , \ \text{else} \end{cases} \ .$$

It suffices to show

$$\text{for } u > v: \ \sum_{w=1}^{k} C_{uw}B_{wv} = 0 \ ,$$

$$\text{for } u \leq v: \ \sum_{w=1}^{k} C_{uw}B_{wv} = \begin{cases} A_{0,l}Q_m & , \ u = v = 1 \\ A_{v-u-1,l}Q_0R + A_{v-u,l}Q_m & , \ \text{else} \end{cases} \ .$$

Because of the recursion formula (4.7), in the later case we have

$$\sum_{w=1}^{k} C_{uw}B_{wv} = A_{v-u,l+1} \ .$$

□

Theorem 4.2. With the notation of (4.4), (4.6) and Definition 4.2 the distribution of $N_t(R)$ is given by

$$P(N_t(R) = k) = \sum_{l=k}^{\infty} \frac{t^l}{l!} \, a' A_{k,l} \mathbf{1} \ , \quad t \geq 0 \ , \ k \in \mathbb{N}_0 \ . \tag{4.8}$$

Proof: First we extend $Q^{(k)}$ to the $m(k+1) \times m(k+1)$ matrix

$$\mathcal{Q}^{(k)} = \begin{pmatrix} Q^{(k)} & 0 \\ 0 & 0 \end{pmatrix} .$$

Now

$$\left(\mathcal{Q}^{(k)}\right)^l = \begin{pmatrix} \left(Q^{(k)}\right)^l & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad (a^{(k)})' Q^{(k)} \mathbf{1} = (a^{(k+1)})' \mathcal{Q}^{(k)} \mathbf{1} .$$

Using the power series expansion in (4.4) we obtain

$$\begin{aligned} \mathrm{P}(N_t(R) = k) &= (a^{(k+1)})' \sum_{l=0}^{\infty} \frac{t^l}{l!} \left(Q^{(k+1)}\right)^l \mathbf{1} - (a^{(k)})' \sum_{l=0}^{\infty} \frac{t^l}{l!} \left(Q^{(k)}\right)^l \mathbf{1} \\ &= \sum_{l=0}^{\infty} \frac{t^l}{l!} (a^{(k+1)})' \left(\left(Q^{(k+1)}\right)^l - \left(\mathcal{Q}^{(k)}\right)^l \right) \mathbf{1} . \end{aligned}$$

For the difference of the matrix powers holds

$$\begin{aligned} \left(Q^{(k+1)}\right)^l - \left(\mathcal{Q}^{(k)}\right)^l = & \begin{pmatrix} A_{0,l} & A_{1,l} & \cdots & A_{k-1,l} & A_{k,l} \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & A_{1,l} & \vdots \\ \vdots & & \ddots & A_{0,l} & A_{1,l} \\ 0 & \cdots & \cdots & 0 & A_{0,l} \end{pmatrix} \\ & - \begin{pmatrix} A_{0,l} & A_{1,l} & \cdots & A_{k-1,l} & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & A_{1,l} & \vdots \\ \vdots & & \ddots & A_{0,l} & \vdots \\ 0 & \cdots & \cdots & 0 & 0 \end{pmatrix} = \left(\begin{array}{c|c} 0 & \begin{matrix} A_{k,l} \\ \vdots \\ A_{0,l} \end{matrix} \end{array}\right) . \end{aligned}$$

Now

$$(a^{(k+1)})' \left(\left(Q^{(k+1)}\right)^l - \left(\mathcal{Q}^{(k)}\right)^l \right) \mathbf{1} = a' A_{k,l} \mathbf{1} ,$$

with $A_{k,l} = 0$ for $l < k$. □

Using the recursion formula (4.7) we may now approximate the distribution of $N_t(R)$, because the summands in (4.8) tend to zero for $l \to \infty$.[3]

[3]For any matrix norm $\|A_{k,l}\| \leq \binom{l}{k} (\max\{\|Q_m\|, \|Q_0 R\|\})^l$. And so

$$\left\| \frac{t^l}{l!} A_{k,l} \right\| \leq \frac{(t \max\{\|Q_m\|, \|Q_0 R\|\})^l}{k!(l-k)!} \ \to \ 0 \ , \text{ for } l \to \infty .$$

Remark 4.5. If the matrices Q_m and Q_0R commutate we have the much simpler form

$$A_{k,l} = \binom{l}{k} Q_m^{l-k}(Q_0R)^k$$

and therefore

$$\mathrm{P}(N_t(R) = k) = a' \sum_{l=k}^{\infty} \frac{t^l}{l!} \binom{l}{k} Q_m^{l-k}(Q_0R)^k \mathbf{1} = \frac{t^k}{k!} a'(Q_0R)^k e^{tQ_m}\mathbf{1} \ .$$

Although this result seems promising, usually the matrices do not commutate.

The mean number of failures $\mu(t, R)$ may be derived without the SSM representation (4.2). For the case that $(\tau_k^R)_{k\in\mathbb{N}}$ is a renewal process with irreducible intensity matrix Ψ^R Theorem 2.4.1 from Neuts [47] states, that

$$\mu(t, R) = \frac{t}{\mathrm{E}(\tau_1)} + \frac{\mathrm{Var}(\tau_1) + (\mathrm{E}(\tau_1))^2}{2\mathrm{E}(\tau_1)} + \frac{1}{\mathrm{E}(\tau_1)}\ a'\left(\Pi - e^{t\Psi^R}\right) Q_m^{-1}\mathbf{I} \ ,$$

where the rows of Π are the unique steady-state distribution of Ψ^R. Using the approach of the proof of this theorem for an arbitrary initial distribution a and intensity matrix Ψ^R, we may obtain again the result of Corollary 4.1.

4.4 Optimality criterion

For a repair model $(M, \mathcal{R}(M))$ we want to choose a repair matrix $R \in \mathcal{R}(M)$ that is optimal in the following sense. Assume that a working system gives some reward like e.g. engine power which is measured by monetary income. The income cumulates due to rates $d_1, ..., d_m > 0$ which depend on the state of the system and we write $d = (d_1, ..., d_m)'$. In case of a failure type i the repair action $r^{i,j} \in \mathcal{R}_i = \{r^{i,1}, ..., r^{i,n_i}\}$ causes costs $c_{ij} \geq 0$, $i = 1, ..., n-m$, $j = 1, ..., n_i$. We denote the costs by the cost matrix $C = (c_{ij})_{i=1,...,n-m, j=1,...,n_{\max}}$, where $n_{\max} = \max\{n_1, ..., n_{n-m}\}$ and the entries c_{ij} that do not correspond to a repair action is defined as zero. The cumulated income up to time T equals

$$I(T, R) = \int_0^T \sum_{j=1}^{m} d_j \mathbb{1}_{\{j\}}(S_u^R)\mathrm{d}u \ ,$$

whereas the cumulated losses up to time T equal

$$L(T, R) = \sum_{i=m+1}^{n} c_{ij_i} N_T(R, i) \ .$$

We call

$$G(T,R) = I(T,R) - L(T,R) \tag{4.9}$$

the cumulated gains up to time T. The expected gains $\mu^G(T,R) = \mathrm{E}(G(T,R))$ up to time $T > 0$ are with Corollary 4.1

$$\begin{aligned}\mu^G(T,R) &= \mathrm{E}(I(T,R) - L(T,R)) \\ &= \int_0^T \sum_{j=1}^m d_j \mathrm{P}(S_u^R = j)\mathrm{d}u - \sum_{i=m+1}^n c_{ir_i}\mu(T,R,i) \\ &= \int_0^T \sum_{j=1}^m \mathrm{P}(S_u^R = j)\left[d_j - \sum_{i=m+1}^n c_{ij_i}q_{ji}\right]\mathrm{d}u \ .\end{aligned} \tag{4.10}$$

Definition 4.6. We call

$$g_j^R := d_j - \sum_{i=m+1}^n c_{ij_i}q_{ji} \tag{4.11}$$

the *gain rate* of state j, $j = 1, ..., m$, and we write $g^R = (g_1^R, ..., g_m^R)'$.

With Theorem 4.1 we obtain

$$\mu^G(T,R) = \int_0^T a' e^{u\Psi^R} g^R \mathrm{d}u \ . \tag{4.12}$$

Definition 4.7. We call a repair matrix $R^* \in \mathcal{R}(M)$

i) optimal up to time $T > 0$, if

$$\mu^G(T,R^*) \geq \mu^G(T,R) \ \forall \ R \in \mathcal{R}(M) \ .$$

ii) finally optimal, if

$$\lim_{T\to\infty} T^{-1}\mu^G(T,R^*) \geq \lim_{T\to\infty} T^{-1}\mu^G(T,R) \ \forall \ R \in \mathcal{R}(M) \ .$$

Remark 4.6. Of course, if there exists a repair matrix R^* with

$$\mu^G(T,R^*) \geq \mu^G(T,R) \ \forall \ R \in \mathcal{R}(M), \ \forall \ T > 0 \ ,$$

then R^* is also finally optimal.

4.5 Optimal repair matrices

We consider a repair model $(M, \mathcal{R}(M))$, reward rates $d_1, ..., d_m$ and respective costs for a repair action c_{ij} , $j = 1, ..., n_i$, $i = 1, ..., n - m$. The gain rates (4.11) are obtained with small effort, hence the main part of the calculation of the expected gains $\mu^G(T, R)$ from (4.10) is to determine the state probabilities of the MC S from Definition 4.2. We investigate four ways how to choose a repair matrix and we will illustrate the results by means of the following example.

Example 4.6. We consider again the failure model $M^{(4)}$ from Example 3.1.

i) The possible repair matrices are given by the sets $\mathcal{R}_1, ..., \mathcal{R}_5$ from Example 4.1. We assume reward rates $d = (d_1, d_2, d_3)' = (100, 80, 50)'$ and repair costs

$$C = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{pmatrix} .$$

ii) The possible repair matrices are all determined repair matrices of dimension 5×3 from Example 4.2. We assume constant reward rates $d = \mathbf{1}$ and repair costs

$$C = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \\ 3 & 2 & 1 \\ 5 & 4 & 3 \end{pmatrix} .$$

4.5.1 Optimal in $[0, T]$

The first approach is a global and therefore exact one. For a given time horizon T we may calculate the expected gains for every possible repair matrix and choose the repair matrix with the highest expected gains. The disadvantage is the computational effort you need for calculating matrix exponentials and the fact that the number of possible repair matrices

$$|\mathcal{R}(M)| = n_1 \cdot ... \cdot n_{n-m}$$

may be large even for models with a small number of working states m.
In order to simplify (4.12) we use the representation of the matrix exponential $e^{t\Psi^R}$ from (2.8) and obtain

$$\mu^G(T,R) = \int_0^T a' e^{u\Psi^R} g^R \mathrm{d}u = \int_0^T a' \left(\sum_{i=1}^{k} \sum_{\alpha=0}^{\nu_i - 1} \frac{u^\alpha e^{\lambda_i u}}{\alpha!} (\Psi^R - \lambda_i \mathbf{I})^\alpha Z_{\lambda_i} \right) g^R \mathrm{d}u \,,$$

where the eigenvalues λ_i and the eigenprojections Z_{λ_i} belong to the intensity matrix Ψ^R. We just need the integrals

$$J(\lambda_i, \alpha, T, R) := \int_0^T \frac{u^\alpha e^{\lambda_i u}}{\alpha!} \, \mathrm{d}u \quad , \; \alpha = 0, ..., \nu_i - 1 \; , \; i = 1, ..., k \; .$$

The solution is

$$J(\lambda_i, \alpha, T, R) = \begin{cases} \dfrac{T^{\alpha+1}}{(\alpha+1)!} & , \text{ if } \lambda_i = 0 \; , \\ (-\lambda_i)^{-(\alpha+1)} \left(1 - e^{\lambda_i T} \sum_{l=0}^{\alpha} \dfrac{(-\lambda_i T)^l}{l!} \right) & , \text{ else } . \end{cases}$$

Hence we avoid numerical integration and the expected gains are given by

$$\mu^G(T,R) = a' \left(\sum_{i=1}^{k} \sum_{\alpha=0}^{\nu_i - 1} J(\lambda_i, \alpha, T, R) (\Psi^R - \lambda_i \mathbf{I})^\alpha Z_{\lambda_i} \right) g^R \; . \tag{4.13}$$

Example 4.7. We consider the repair models from Example 4.6 and calculate the expected gains up to T for all possible repair matrices.

i) For the 16 possible repair matrices from Example 4.6 i) we calculate the expected gains (4.13). Figure 4.1 shows the functions $\mu^G(T,R)/T$ for $T \in [0,1]$ respectively $T \in [0,25]$. The functions in Figure 4.1 look smooth, they seem to converge to a constant. We evaluate the expected gains for $T = 1$ and $T = 25$ (Figure 4.2). The number of a repair matrix $R = (r^{1,j_1}|...|r^{n-m,j_{n-m}})'$ is given by the lexicographical order of the index vector $j = (j_1, ..., j_{n-m})$. The highest expected gains result from repair matrix number 16, which corresponds to the index vector $j = (1,2,2,2,2)$ (cf. Example 4.1). This is the repair matrix

$$R_{16} = \begin{pmatrix} 1 & 0 & 0 \\ 0.8 & 0.2 & 0 \\ 0.6 & 0.3 & 0.1 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.6 & 0.2 \end{pmatrix} \; .$$

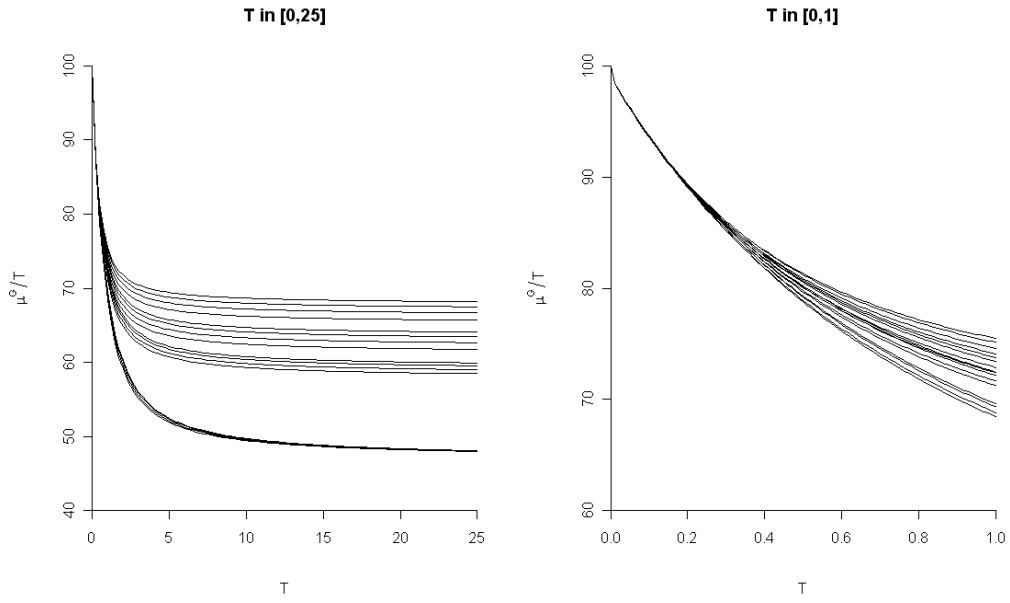


Figure 4.1: Expected gains for the possible repair matrices in Example 4.6 i)

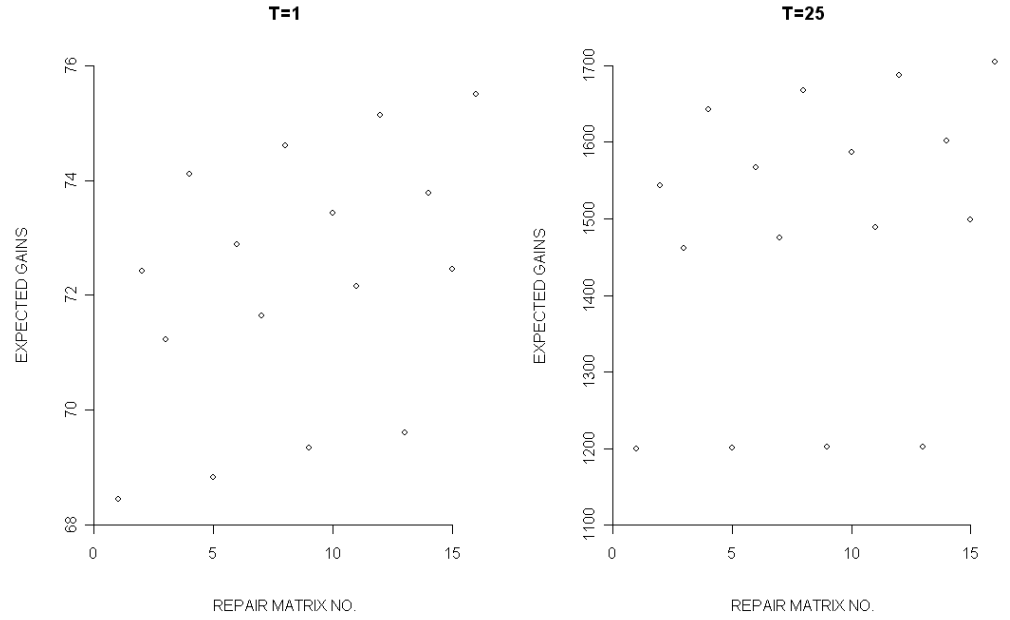


Figure 4.2: Expected gains at $T = 1$ and $T = 25$ for Example 4.6 i)

ii) For $T = 1$ and $T = 25$ we calculate the expected gains (4.13) for the $3^5 = 243$ possible determined repair matrices from Example 4.6 ii).

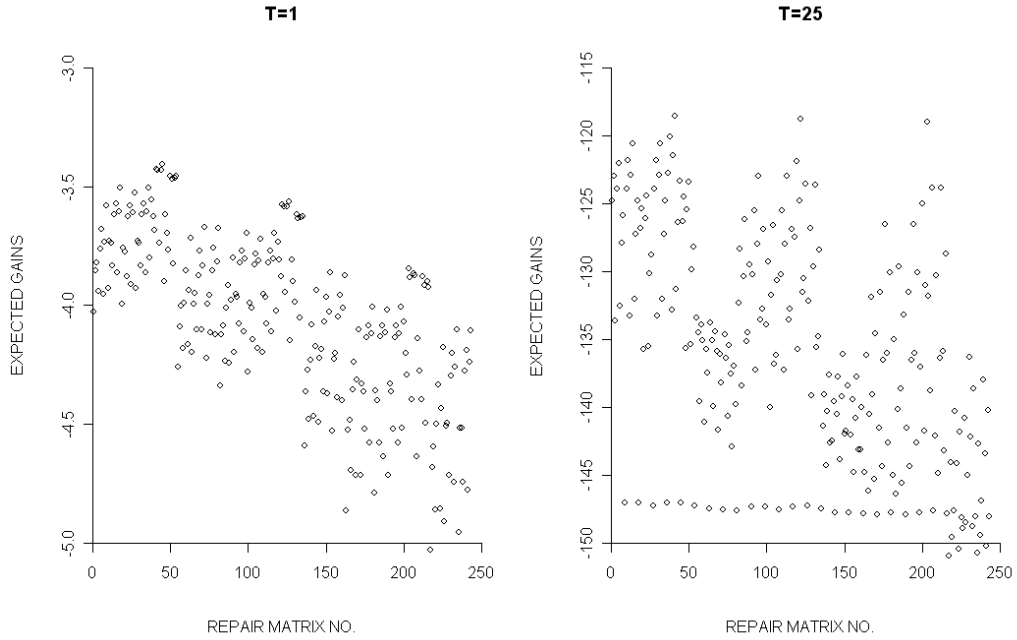


Figure 4.3: Expected gains at $T = 1$ and $T = 25$ for Example 4.6 ii)

For $T = 1$ the eight repair matrices with the highest expected gains are

no.	45	41	42	44	50	54	53	51
μ^G	-3.41	-3.43	-3.43	-3.43	-3.46	-3.46	-3.47	-3.47

.

For $T = 25$ the seven repair matrices with the highest expected gains are

no.	41	122	203	38	14	32	40
μ^G	-118.6	-118.8	-119.0	-120.1	-120.6	-120.6	-121.5

.

The best repair matrix for $T = 1$ is R_{45}. But for $T = 25$ this matrix yields expected gains of $\mu^G(25, R_{45}) = -147$ which is not even close to maximal expected gains. The repair matrix R_{41} looks promising for small and large values of T. The corresponding index vector equals

$J = (1, 2, 2, 2, 2)$ which gives the repair matrix

$$R_{41} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} .$$

4.5.2 Finally optimal

Considering the ratio

$$\frac{\mu^G(T,R)}{T} = \sum_{j=1}^{m} g_j^R \frac{1}{T} \int_0^T \mathrm{P}(S_u^R = j)\mathrm{d}u \ , \ T > 0 \tag{4.14}$$

we notice that the integrands converge to the components of the steady-state distribution $x^{a,R}$ of S^R (cf. Lemma 2.5):

$$\lim_{u\to\infty} \mathrm{P}(S_u^R = j) = x_j^{a,R} \ \ , \ j = 1, ..., m \ . \tag{4.15}$$

Lemma 4.4. The ratio $\dfrac{\mu^G(T,R)}{T}$ from (4.14) converges for $T \to \infty$:

$$\lim_{T\to\infty} \frac{\mu^G(T,R)}{T} = \sum_{j=1}^{m} g_j^R x_j^{a,R} = (x^{a,R})' g^R \ .$$

Proof: Since (4.15) for all $\delta > 0$ exists $u_\delta > 0$ with

$$|\mathrm{P}(S_{u_\delta}^R = j) - x_j^{a,R}| \leq \delta \ \ , \ \forall \ j = 1, ..., m \ .$$

Now for $T > u_\delta$ holds

$$\begin{aligned} &\left| \sum_{j=1}^{m} g_j^R \frac{1}{T} \int_0^T \mathrm{P}(S_u^R = j)\mathrm{d}u - \sum_{j=1}^{m} g_j^R x_j^{a,R} \right| \\ &\leq \sum_{j=1}^{m} |g_j^R| \frac{1}{T} \int_0^T |\mathrm{P}(S_u^R = j) - x_j^{a,R}|\mathrm{d}u \\ &\leq \sum_{j=1}^{m} |g_j^R| \left(\frac{1}{T} \int_0^{u_\delta} |\mathrm{P}(S_u^R = j) - x_j^{a,R}|\mathrm{d}u + \delta \right) =: \epsilon(T,\delta) \ . \end{aligned}$$

If we take T large enough and δ small enough, $\epsilon(T,\delta)$ is arbitrarily close to zero. □

Remark 4.7. For an irreducible MC with intensity matrix Q the rate of convergence to the steady-state distribution equals $e^{-\rho t}$, where $\rho = -\mathrm{Re}(\lambda)$ and λ is the non-zero eigenvalue of Q with largest real part (cf. Kijima [34] Section 4.5 and note that all non-zero eigenvalues of Q have a negative real part).

Now we just calculate this limit for every possible repair matrix.

Example 4.8. We consider the repair models from Example 4.6. The limiting expected gains per time unit $(x^{a,R})'g^R$ from Lemma 4.4 for the possible repair functions are shown in the next figure

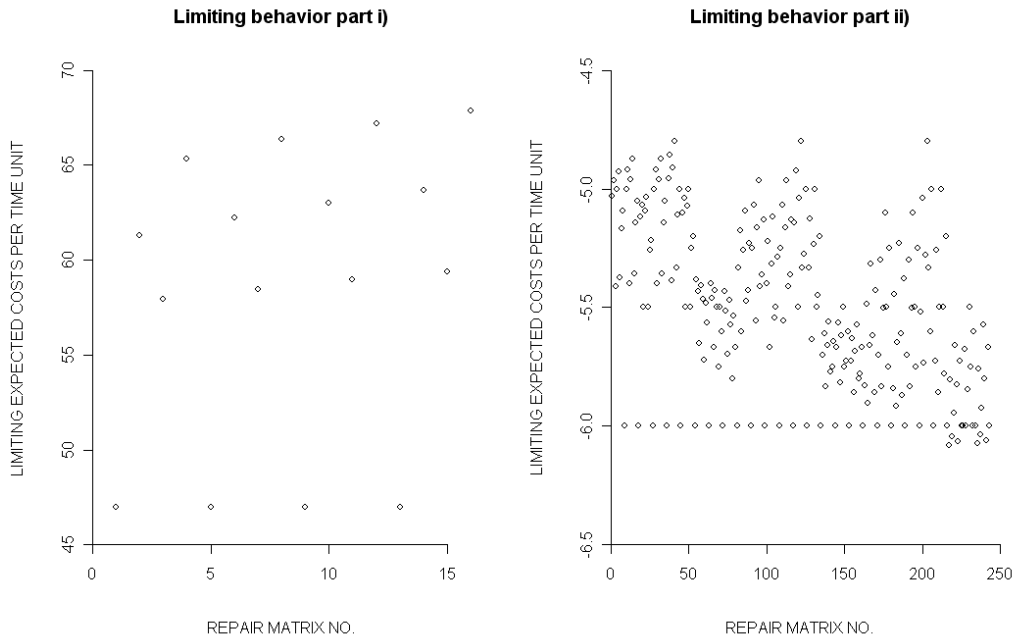


Figure 4.4: Limiting expected gains per time unit for Example 4.6

i) The repair matrix R_{16} is finally optimal with $(x^{a,R_{16}})'g^{R_{16}} = 67.88$.

ii) The three repair matrices R_{41}, R_{122} and R_{203} are finally optimal with $(x^{a,R})'g^R = -4.8$. We know the first repair matrix from Example 4.7

and the other two are given by

$$R_{122} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \ , \ R_{203} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \ .$$

Remark 4.8. The limiting distribution can also be obtained in a different way. Remember that for a given repair matrix R holds (4.15)

$$\lim_{t\to\infty} \mathrm{P}(S_t^R = j) = \lim_{t\to\infty} a' e^{t\Psi^R} e_j = x_j^{a,R} \ .$$

If we apply (2.8)

$$x_j^{a,R} = \lim_{t\to\infty} a' \left(\sum_{i=1}^{k} \sum_{\alpha=0}^{\nu_i - 1} \frac{t^\alpha e^{\lambda_i t}}{\alpha!} (\Psi^R - \lambda_i \mathbf{I})^\alpha Z_{\lambda_i} \right) e_j$$

we see that all summands vanish except for $\lambda_i = 0$. If $\nu = \mathrm{ind}(\Psi^R)$ is the index of the eigenvalue $\lambda = 0$, we have

$$x_j^{a,R} = \lim_{t\to\infty} a' \left(\sum_{\alpha=0}^{\nu-1} \frac{t^\alpha}{\alpha!} (\Psi^R)^\alpha Z_0 \right) e_j \ .$$

Since $x_j^{a,R}$ is bounded and the right-hand side is the limit of a polynomial, this polynomial has to be constant. This shows

$$x_j^{a,R} = a' Z_0 e_j \quad , \ j = 1, ..., m \ .$$

Especially, the index of Ψ^R (and of any other intensity matrix) equals one. The disadvantage of this approach is, that we need to calculate the eigenvalues of Ψ^R, which may again cause numerical instability (cf. Remark 2.8).

4.5.3 Empirical approach

Next we want to avoid calculating the distribution of S. First, we simulate the processes $N(R, i)$, $i = m+1, ..., n$ in order to obtain estimates $\widehat{\mu}(T, R, i)$. Let $N_1, ..., N_k$ be an iid sample with $\mathrm{P}^{N_1} = \mathrm{P}^{N_T(R,i)}$. A natural candidate as an estimator for $\mu(T, R, i)$ is

$$\widehat{\mu}(T, R, i) := \frac{1}{k} \sum_{j=1}^{k} N_j \ .$$

Additionally, we can estimate covariances between the numbers of distinct failure types if we observe these numbers simultaneously: Let $\mathbf{N}^1, ..., \mathbf{N}^k$ be an iid sample of random vectors whose joint distribution $\mathrm{P}^{\mathbf{N}^1}$ equals the joint distribution $\mathrm{P}^{(N_T(R,m+1),...,N_T(R,n))}$. Now the canonical estimators for the mean vector $\mu(T,R) = (\mu(T,R,m+1), ..., \mu(T,R,n))'$ and the covariance matrix $\Sigma(T,R) = (\mathrm{Cov}(N_T(R,i+m), N_T(R,j+m)))_{i,j=1,...,n-m}$ are

$$\widehat{\mu}_i(T,R) := \frac{1}{k}\sum_{l=1}^{k} \mathbf{N}_i^l \qquad , \ i = 1, ..., n-m \ ,$$

$$\widehat{\Sigma}_{ii}(T,R) := \frac{1}{k-1}\sum_{l=1}^{k} (\mathbf{N}_i^l - \widehat{\mu}_i(T,R))^2 \qquad , \ i = 1, ..., n-m \ ,$$

$$\widehat{\Sigma}_{ij}(T,R) := \frac{1}{k-2}\sum_{l=1}^{k} (\mathbf{N}_i^l - \widehat{\mu}_i(T,R))(\mathbf{N}_j^l - \widehat{\mu}_j(T,R)) \qquad , \ i \neq j \ .$$

Neuts and Pagano [49] used the alias algorithm (Section 2.7) for generating PH-distributed random variables. In detail you apply the alias algorithm for the rows of the one-step transition matrix (2.10) of the MCs $(J^{(k)})_{k\in\mathbb{N}}$ from Definition 4.2. The times between the jumps are exponentially distributed. In a similar fashion we may also simulate the path of S, so we may in fact simulate the cumulated gain $G(T,R)$ from (4.9). An iid sample $G_1, ..., G_k$ with $\mathrm{P}^{G_1} = \mathrm{P}^{G(T,R)}$ leads to the estimators

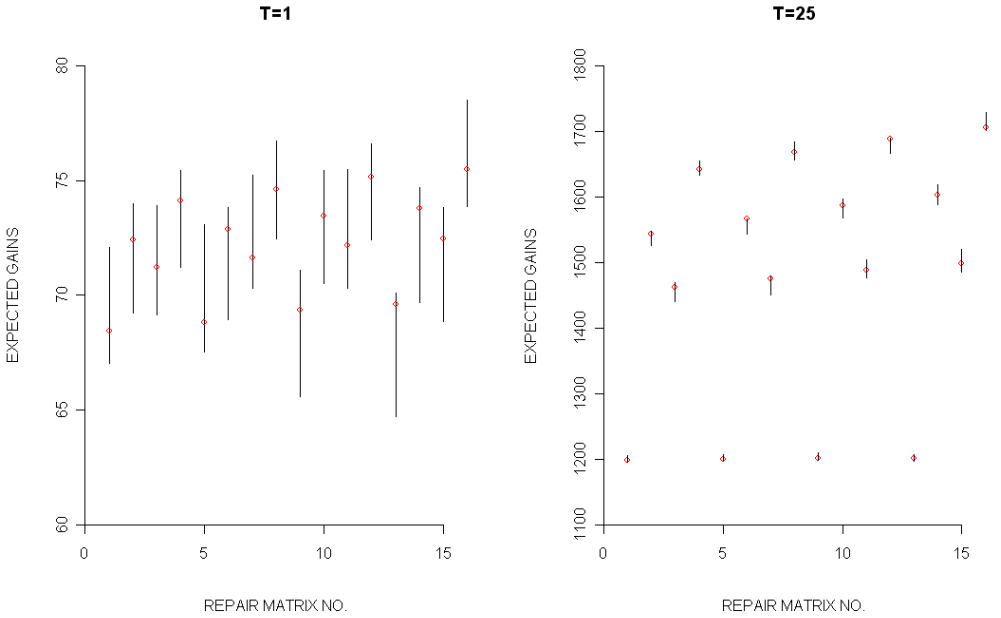
$$\widehat{\mu^G}(T,R) := \frac{1}{k}\sum_{l=1}^{k} G_l \ , \ \widehat{\Sigma^G}(T,R) := \frac{1}{k-1}\sum_{l=1}^{k} (G_l - \widehat{\mu^G}(T,R))^2 \ ,$$

for $\mu^G(T,R)$ and $\Sigma^G(T,R) := \mathrm{Var}(G(T,R))$. If the sample size k is large enough we may use standard statistical tools to confirm whether a repair matrix significantly leads to lower expected costs than another repair matrix.

Example 4.9. We consider the repair models from Example 4.6. For every possible repair matrix we simulate $k = 100$ cumulated gains $G(T,R)$ for $T = 1$ and for $T = 25$. We calculate the estimates for the parameters $\mu^G(T,R)$ and $\Sigma^G(T,R)$ and calculate corresponding approximate 95% confidence intervals, using the interval estimators

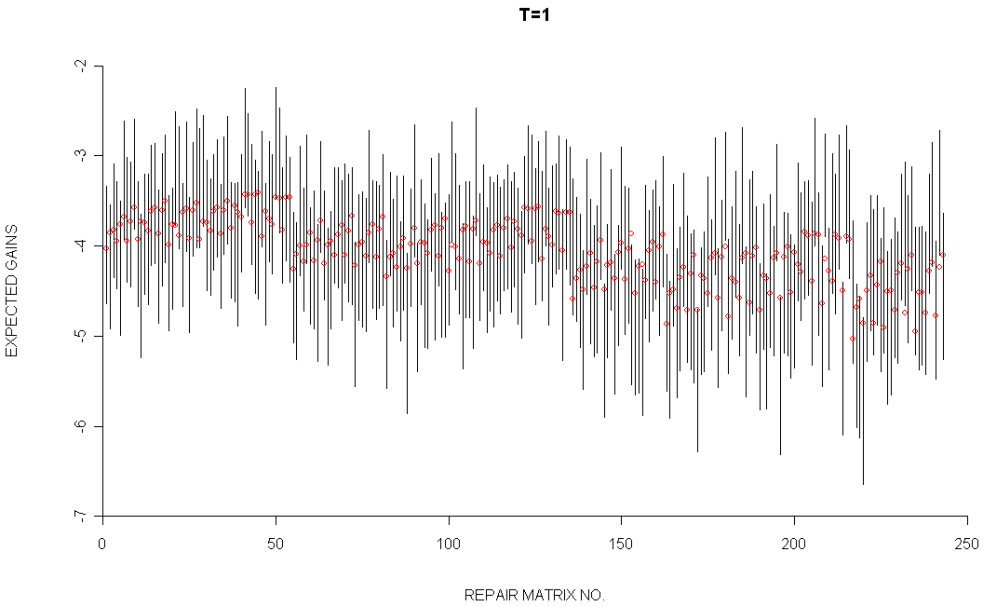
$$\widehat{I}_{\mu^G(T,R)} = \left[\widehat{\mu^G}(T,R) - 1.96\sqrt{\frac{\widehat{\Sigma^G}(T,R)}{k}}, \widehat{\mu^G}(T,R) + 1.96\sqrt{\frac{\widehat{\Sigma^G}(T,R)}{k}}\right] \ .$$

We plot the confidence intervals and the true values which we calculated in Example 4.7.



i)

Figure 4.5: Confidence intervals for the expected gains for Example 4.6 i)



ii)

Figure 4.6: Confidence intervals for the expected gains for Example 4.6 ii), $T = 1$

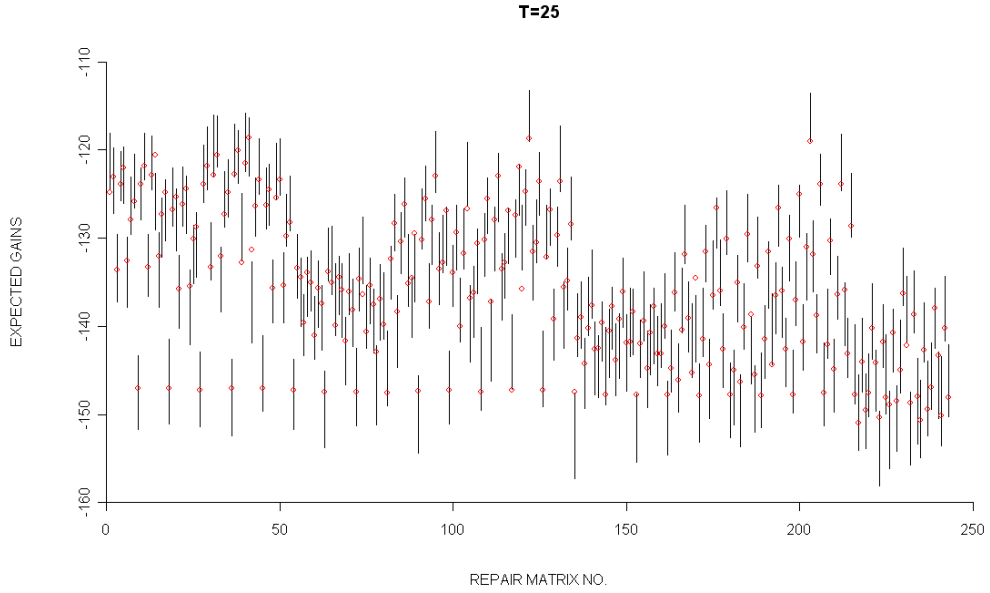


Figure 4.7: Confidence intervals for the expected gains for Example 4.6 ii), $T = 25$

4.5.4 Heuristic approach

The three first approaches have in common that we investigate all possible repair matrices. Now we propose a heuristic approach, which investigates only few repair matrices and which shall result in a good repair matrix. In case of a failure of type j, $j \in \{m+1, ..., n\}$ one has to decide which repair action from $\mathcal{R}_{j-m}$ to perform. The repair action $r^{j-m,l} \in \mathcal{R}_{j-m}$ causes immediate costs $c_{j-m,l}$, $l = 1, ..., n_{j-m}$, and until the next failure happens income is yielded.

Lemma 4.5. Consider the failure model $M = (a, Q)$. Let the system be in state $i \in \{1, ..., m\}$. The expected income until the next failure μ_i^I equals

$$\mu_i^I = -e_i' Q_m^{-1} D\mathbf{1} \ ,^4$$

where $D = \text{diag}(d_1, ..., d_m)$. We define $\mu^I := (\mu_1^I, ..., \mu_m^I)'$.

[4]This equals the expectation of a bilateral phase-type distribution (cf. Ahn and Ramaswami [5]) with representation (e_i, Q_m, D).

Proof: Remember the processes $J^{(1)}$, S and (τ^R, U^R) from Definition 4.2 and assume that $S_0^R = i$. Now the income until the next failure is

$$\int_0^{\tau_1^R} \sum_{j=1}^{m} d_j \mathbb{1}_{\{j\}}(S_u^R) \mathrm{d}u \ .$$

Since the MC $J^{(1)}$ with $a = e_i$ equals S^R on $[0, \tau_1^R)$ and takes values in $\{m+1, ..., n\}$ on $[\tau_1^R, \infty)$ the income may be written as

$$\int_0^{\infty} \sum_{j=1}^{m} d_j \mathbb{1}_{\{j\}}(J_u) \mathrm{d}u \ .$$

Taking conditional expectations yields

$$\begin{aligned} \mu_i^I &= \int_0^{\infty} \sum_{j=1}^{m} d_j \mathrm{P}(J_u^{(1)} = j | J_0^{(1)} = i) \mathrm{d}u = \int_0^{\infty} \sum_{j=1}^{m} d_j e_i' e^{Q_m u} e_j \mathrm{d}u \\ &= \int_0^{\infty} e_i' e^{Q_m u} D\mathbf{1} \mathrm{d}u = -e_i' Q_m^{-1} D\mathbf{1} \ . \end{aligned}$$

□

Remember that in Lemma 2.3 we proved that

$$\mu_i^\tau = \mathrm{E}(\tau_1^R | J_0^{(1)} = i) = -e_i' Q_m^{-1} \mathbf{1} \ , \ 1 = 1, ..., m \ .$$

We write $\mu^\tau = (\mu_1^\tau, ..., \mu_m^\tau)'$. A decision rule, which may be interpreted as maximizing the gain rate until the next failure, is to maximize the ratio of the expected gains per expected time until the next failure:

$$\forall \ j \in \{m+1, ..., n\} \ \underset{l=1,...,n_{j-m}}{\text{maximize}} \ \frac{(r^{j-m,l})' \mu^I - c_{j-m,l}}{(r^{j-m,l})' \mu^\tau} \ . \tag{4.16}$$

But this decision rule alone does not cope with the complexity of the model. In the following example we see that maximizing the gain rates might cause the system to evolve to more severe states.

Example 4.10. Let the system describe a mobile phone. If the mobile phone is new, a failure may occur when the battery breaks down. Now we may either buy a new battery from the original manufacturer, or we may buy a low-budget battery from another manufacturer. Assume that a low-budget battery does not simply break down, but implodes and destroys the mobile phone. In the latter case we may either buy a new mobile phone with an

original battery or we buy a new mobile phone with a low-budget battery. Let the parameters of the model be $m = 2$, $n = 4$, $a = e_1$ and

$$Q = \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \ , \ \mathcal{R}_1 = \mathcal{R}_2 = \{e_1, e_2\} \ , \ C = \begin{pmatrix} 3 & 1 \\ 10 & 8 \end{pmatrix} \ , \ d = \mathbf{1} \ .$$

If we apply rule (4.16) with $\mu^I = -Q_2^{-1} D\mathbf{1} = \mathbf{1}$ and $\mu^\tau = -Q_2^{-1}\mathbf{1} = \mathbf{1}$ we obtain

$$\begin{aligned} j = 3: \quad l = 1: \quad & \frac{e_1' \mu^I - c_{11}}{e_1' \mu^\tau} = \frac{1-3}{1} = -2 \ , \\ l = 2: \quad & \frac{e_2' \mu^I - c_{12}}{e_2' \mu^\tau} = \frac{1-1}{1} = 0 \ , \\ j = 4: \quad l = 1: \quad & \frac{e_1' \mu^I - c_{21}}{e_1' \mu^\tau} = \frac{1-10}{1} = -9 \ , \\ l = 2: \quad & \frac{e_2' \mu^I - c_{22}}{e_2' \mu^\tau} = \frac{1-8}{1} = -7 \ . \end{aligned}$$

This results in the repair matrix $R_1 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$. Whether this is a good repair matrix may be calculated with Example 2.6 and Lemma 4.4. We have

$$\begin{aligned} & \Psi^{R_1} = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \ , \ e^{t\Psi^{R_1}} = \begin{pmatrix} e^{-t} & 1 - e^{-t} \\ 0 & 1 \end{pmatrix} \ , \ x^{a,R} = e_2 \ , \ g_2^R = -7 \\ & \Rightarrow \lim_{T \to \infty} \frac{\mu^G(T, R_1)}{T} = -7 \ . \end{aligned}$$

We compare R_1 with the repair matrix $R_2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$.

$$\begin{aligned} & \Psi^{R_2} = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix} \ , \ e^{t\Psi^{R_2}} = \begin{pmatrix} 1 & 0 \\ 1 - e^{-t} & e^{-t} \end{pmatrix} \ , \ x^{a,R} = e_1 \ , \ g_1^R = -2 \\ & \Rightarrow \lim_{T \to \infty} \frac{\mu^G(T, R_1)}{T} = -2 \ . \end{aligned}$$

This shows that R_2 yields larger gains per time than R_1 (In fact, R_2 is finally optimal.).

We need to take into account forthcoming failures. The approach is that we use rule (4.16) to determine the repair matrix that we apply for a virtual

last failure and then determine the repair matrix which is optimal for the next to last failure and so on. If we restart after the next to last failure in state $i \in \{1,...,m\}$ the probability that the last failure is of type $j+m$, $j \in \{1,...,n-m\}$ equals b_{ij} from Lemma 2.2

$$b_{ij} = \mathrm{P}(J_\tau = j+m | J_0 = i) \;\Rightarrow\; B = -Q_m^{-1} Q_0 \ .$$

The decision rule may now be interpreted as cumulated gain rate. We use the following algorithm:

(Step 1) for each failure type $j = m+1,...,n$ do:

$$\begin{aligned} l^* &:= \underset{l=1,...,n_{j-m}}{\operatorname{argmax}} \frac{(r^{j-m,l})'\mu^I - c_{j-m,l}}{(r^{j-m,l})'\mu^\tau}, \\ r_{j-m}^{(1)} &:= l^*, \\ G_{j-m}^{(1)} &:= (r^{j-m,l^*})'\mu^I - c_{j-m,l^*}, \\ T_{j-m}^{(1)} &:= (r^{j-m,l^*})'\mu^\tau. \end{aligned}$$

(Step s) $s \geq 2$, for each failure type $j = m+1,...,n$ do:

$$\begin{aligned} l^* &:= \underset{l=1,...,n_{j-m}}{\operatorname{argmax}} \frac{(r^{j-m,l})'\mu^I - c_{j-m,l} + (r^{j-m,l})'BG^{(s-1)}}{(r^{j-m,l})'\mu^\tau + (r^{j-m,l})'BT^{(s-1)}}, \\ r_{j-m}^{(s)} &:= l^*, \\ G_{j-m}^{(s)} &:= (r^{j-m,l^*})'\mu^I - c_{j-m,l^*} + (r^{j-m,l^*})'BG^{(s-1)}, \\ T_{j-m}^{(s)} &:= (r^{j-m,l^*})'\mu^\tau + (r^{j-m,l^*})'BT^{(s-1)}. \end{aligned}$$

Here $G^{(s)} = (G_1^{(s)},...,G_{n-m}^{(s)})'$ and $T^{(s)} = (T_1^{(s)},...,T_{n-m}^{(s)})'$ are s-failure gains respective s-failure times given the current failure type, and $r^{(s)} = (r_1^{(s)},...,r_{n-m}^{(s)})'$ gives the current repair matrix $R^{(s)}$, where the i-th row of $R^{(s)}$ is given by $r^{i,r_i^{(s)}} \in \mathcal{R}_i$, $i = 1,...,n-m$ from Definition 4.1.

(STOP) Stop if for a fixed integer $K \geq 1$

$$\exists\ s_1,...,s_K :\ 1 \leq s_1 < ... < s_K,\ r^{(s_1)} = ... = r^{(s_K)} \ .$$

Then we choose the repair matrix $R^{(s_K)}$.

According to this algorithm, we choose the first repair matrix that is optimal in K iterations of the algorithm. In the next example we show that the result of the algorithm may depend on the parameter K.

Example 4.11. Consider the repair model with $m \geq 2$ fixed, $a = e_1$ and

$$\begin{aligned} &Q = (-\mathbf{I} \mid \mathbf{I}) \in \mathbb{R}^{m \times 2m} \ , \\ &\mathcal{R}_1 = \{e_1, e_2\} \ , \\ &\mathcal{R}_m = \{e_m\} \ , \\ &\mathcal{R}_i = \{e_{i+1}\} \ , \ i = 2, ..., m-1 \ , \\ &d = \mathbf{1} \ , \end{aligned} \qquad C = \begin{pmatrix} 2 & 1 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 3 & 0 \end{pmatrix} .^{5}$$

There are only two possible repair matrices. We calculate which one of them is finally optimal.

$$R_1 = \begin{pmatrix} e_1' \\ e_3' \\ \vdots \\ e_m' \\ e_m' \end{pmatrix} , \quad \Psi^{R_1} = \begin{pmatrix} 0 & \cdots & 0 \\ & * & \end{pmatrix} \ \Rightarrow \ x^{a,R} = e_1 \ ,$$
$$g_1^R = 1 - 2 = -1 \ , \ \lim_{T \to \infty} \frac{\mu^G(R_1, T)}{T} = -1 \ .$$

$$R_2 = \begin{pmatrix} e_2' \\ e_3' \\ \vdots \\ e_m' \\ e_m' \end{pmatrix} , \quad \Psi^{R_2} = \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ & & & 0 \end{pmatrix} \ \Rightarrow \ x^{a,R} = e_m \ ,$$
$$g_m^R = 1 - 3 = -2 \ , \ \lim_{T \to \infty} \frac{\mu^G(R_2, T)}{T} = -2 \ .$$

This shows that R_1 is finally optimal. If we use the algorithm presented above, the resulting repair matrices are

$$R^{(s)} = R_2 \ \text{ for } 1 \leq s \leq m-1 \ , \ \text{ and } \ R^{(s)} = R_1 \ \text{ for } m \leq s \ .$$

Consequently, if we use the parameter K with $K < m$, the algorithm yields the worse repair matrix.

The example above motivates us to take $K \geq m$. We should also check the repair matrix $R^{(s_K)}$ that results from the algorithm by calculating the exact gains (either in some interval $[0, T]$ or the limiting behavior) of $R^{(1)}, ..., R^{(s_K)}$. If one of these repair matrices yields gains larger than the gains of $R^{(s_K)}$, we should not accept $R^{(s_K)}$ as best choice.

---

[5] If $m = 2$ we take $C = \begin{pmatrix} 2 & 1 \\ 3 & 0 \end{pmatrix}$, $\mathcal{R}_2 = \{e_2\}$, and if $m = 3$ we take $C = \begin{pmatrix} 2 & 1 \\ 1 & 0 \\ 3 & 0 \end{pmatrix}$.

Remark 4.9. In the algorithm presented above we choose the repair matrix $R \in \mathcal{R}(M)$ row by row. The obtained repair matrix does not need to be applicable with respect to an information level $\mathfrak{F}$. If we use one of the first three approaches, we may detect the applicable repair matrix, and we may only consider the applicable repair matrices for the optimization. This might especially be effective if the number of possible repair matrices is small, and if it is relatively small concerning the possible repair matrices.

Example 4.12. We consider the repair models from Example 4.6 and use the algorithm presented above for choosing a repair matrix. We take the parameter $K = 5$.

i) The steps of the algorithm yield the following repair matrices:

Step	j_1	j_2	j_3	j_4	j_5	no.
1	1	2	2	2	2	16
2	1	2	2	2	2	16
3	1	2	2	2	2	16
4	1	2	2	2	2	16
5	1	2	2	2	2	16

The repair matrix R_{16} is obtained in all steps of the algorithm. According to Example 4.8, the repair matrix R_{16} is finally optimal.

ii) The steps of the algorithm yield the following repair matrices:

Step	j_1	j_2	j_3	j_4	j_5	no.
1	1	2	3	3	3	54
2	1	2	2	2	2	41
3	1	2	2	2	2	41
4	1	2	2	2	2	41
5	1	2	2	2	2	41
6	1	2	2	2	2	41

The resulting repair matrix R_{41} is known to be finally optimal (Example 4.8) and also optimal in $[0, 25]$ (Example 4.7). The repair matrix R_{54} which is obtained in the first step of the algorithm yields good results for short time horizons (Example 4.7).

Chapter 5

Preventive maintenance actions

We want to perform preventive maintenance actions in order to avoid failures of the system. This makes sense if the costs for a preventive maintenance action are small compared to the repair costs and if some working states of the system are preferable, in particular if there exists a state which may be interpreted as an 'as-good-as-new' state. The usual idea is that the system starts in this state and after the preventive maintenance action the system relaunches in this state. We call this approach replacement of the system. A replacement policy is given by a stopping time T with respect to some filtration $\mathfrak{F}$. We assume that we always have a filtration, such that the replacement policy T is a stopping time. Assume that T has a finite but non-zero expectation and let $G(T, R)$ be the cumulated gain up to time T. We want to maximize the expected gains per time unit in the renewal interval $[0, T]$ over suitable classes of stopping times. We present two classical maintenance policies and derive a new policy which is more suitable for our model. Let us first consider an example.

Example 5.1. Let us consider a system with two working states and two failure types ($m = 2$, $n = 4$) with

$$a = e_1 \ , \quad Q_m = \begin{pmatrix} -1.1 & 1 \\ 0 & -1 \end{pmatrix} \ , \quad Q_0 = \begin{pmatrix} 0.1 & 0 \\ 0 & 1 \end{pmatrix} \ , \quad d = 0.$$

State 1 may be interpreted as good condition of the system and state 2 as bad condition. In case of a failure the system is replaced, which means the repair matrix is $R = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$. Repair costs are equal to 1. We calculate

the limiting behavior of the expected gains from Lemma 4.4

$$\Psi^R = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \ , \quad g^R = \begin{pmatrix} -0.1 \\ -1 \end{pmatrix} \ , \quad x^{a,R} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \ ,$$
$$\Rightarrow \lim_{T\to\infty} \frac{\mu^G(T,R)}{T} = (x^{a,R})' g^R = -0.55 \ .$$

The eigenvalues of Ψ^R are $\lambda_1 = 0$ and $\lambda_2 = -2$, both with index one. With (2.7) the eigenprojections are

$$Z_0 = \frac{\Psi^R + 2\mathbf{I}}{2} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} \ , \ Z_{-2} = \frac{\Psi^R + 2\mathbf{I} - 2\mathbf{I}}{-2} = \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix} \ .$$

We apply (4.13) and obtain

$$J(0,0,T,R) = T \ , \ J(-2,0,T,R) = 0.5(1-e^{-2T})$$
$$\mu^G(T,R) = e_1' \left(T \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix} + 0.5(1-e^{-2T}) \begin{pmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{pmatrix} \right) \begin{pmatrix} -0.1 \\ -1 \end{pmatrix}$$
$$= -0.55T + 0.225(1-e^{-2T}) \ .$$

We investigate three inspection policies, where $c > 0$ are the costs for a preventive replacement.

(1) At time T a preventive replacement is performed. The expected gains in the interval $[0,T]$ are $\mu^G(T,R) - c$. The expected gains per time unit in the interval $[0,T]$ are hence

$$g(T,c) = \frac{\mu^G(T,R) - c}{T} = \frac{-22T + 9(1-e^{-2T}) - 40c}{40T}$$

which we want to maximize with respect to T for a fixed $c > 0$. The first order condition is

$$c = \frac{9}{40}(1 - e^{-2T}(1+2T))$$

which has a unique solution $T^*(c)$ iff $0 < c < 9/40$ with

$$g(T^*(c),c) = -11/20 + 9/20 e^{-2T^*(c)} \ .$$

(2) At time T a preventive replacement is performed if no failure happened up to time T. After the first replacement at time $\tau_1^R \wedge T$ the same

rule applies, say a preventive replacement is performed if no failure happened up to time $(\tau_1^R \wedge T) + T$. The expected gains in the interval $[0, \tau_1^R \wedge T]$ are henceforth

$$-\mathrm{P}(\tau_1^R \leq T) - c \cdot \mathrm{P}(\tau_1^R > T) \ .$$

The expected length of a renewal cycle equals

$$\int_0^T u f_{\tau_1^R}(u) \mathrm{d}u + T \cdot \mathrm{P}(\tau_1^R > T) \ .$$

We have to maximize

$$g(T, c) = \frac{(1-c)(10e^{-T} - 9e^{-1.1T}) - 1}{\frac{20}{11} + \frac{90}{11}e^{-1.1T} - 10e^{-T}}$$

with respect to T for a fixed $c > 0$. The first order condition yields

$$c = \frac{9 - 99e^{T} + 90e^{1.1T}}{9 - 198e^{T} + 200e^{1.1T}} \ .$$

The right-hand side is monotone increasing in $T \geq 0$ with values in $[0, 0.45)$. Hence the first order condition has a unique solution $T^*(c)$ for $0 < c < 9/20$ with

$$g(T^*(c), c) = 1.1 \frac{1989 - 198e^{T^*(c)} + 200e^{1.1T^*(c)} - 891e^{-0.1T^*(c)} - 1100e^{0.1T^*(c)}}{(9 - 198e^{T^*(c)} + 200e^{1.1T^*(c)})(2 - 11e^{-T^*(c)} + 9e^{-1.1T^*(c)})} \ .$$

(3) We perform a preventive replacement, if the system runs into state 2 (the bad condition state)

$$T = \inf\{t \geq 0 : S_t^R = 2\} \ .$$

Let N be the number of failures before T. $N + 1$ is geometrically distributed with $p = 10/11$, so $\mathrm{E}(N) = 0.1$. T is the $(N+1)$-fold convolution of exponentially distributed random variables with mean $10/11$. Hence $\mathrm{E}(T) = \frac{10}{11}\mathrm{E}(N+1) = 1$. The expected gains per time in a renewal circle are

$$g(c) = \frac{-\mathrm{E}(N) - c}{\mathrm{E}(T)} = -0.1 - c \ .$$

This policy should only be applied if the expected gains per time are at least equal to the gains per time without preventive maintenance,

$$g(c) \geq -0.55 \ \Leftrightarrow \ c \leq 0.45 \ .$$

We compare the three policies graphically

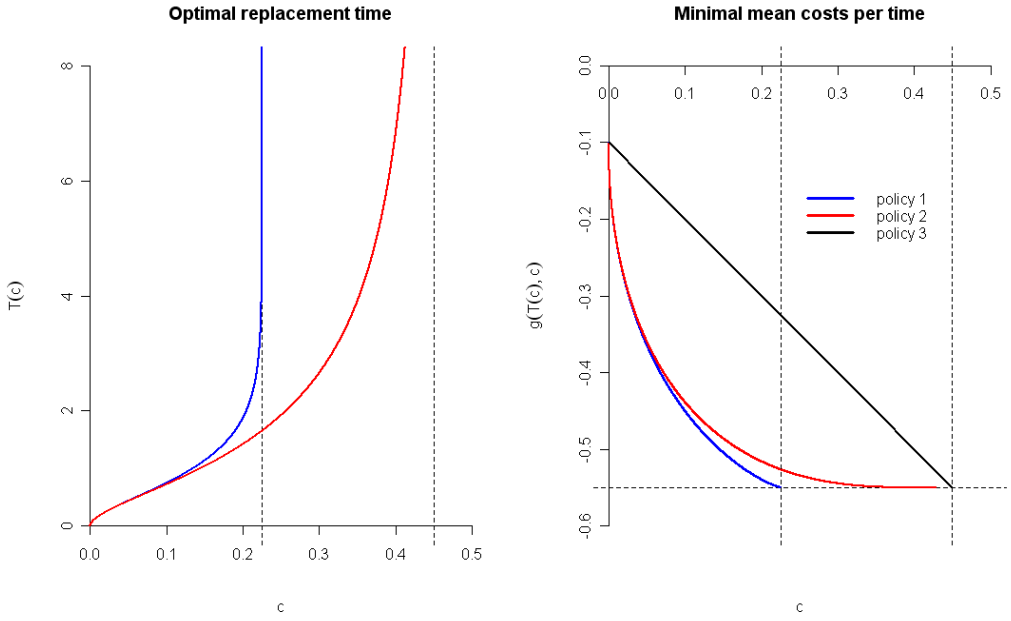


Figure 5.1: Preventive maintenance: three policies

The graphical solution shows, that the third policy clearly outperforms the first two policies if $c < 0.45$.

The first two policies in Example 5.1 are classical preventive maintenance policies. The first policy is a *block replacement policy* and the second policy is an *age replacement policy*. The third policy uses information about the underlying MC S^R, and it seems like, that this advance of knowledge about the proneness to failure of the system helps to obtain larger expected gains than the first two policies. We call the third policy a *critical state maintenance policy*.

5.1 Block replacement policy

Let $(M = (a, Q), \mathcal{R}(M))$ be a repair model. The original approach of the block replacement policy is, that there exists an 'as-good-as-new' state. Assume that state 1 is the 'as-good-as-new' state. We now consider $a = e_1$ and assume that upon failure the system is repaired with respect to the repair

matrix R. A preventive replacement at costs $c > 0$ is performed at a fixed time $T > 0$. The expected gain per time unit is the ratio

$$g(T, c) = \frac{\mu^G(T, R) - c}{T} \ .$$

Clearly, $g(T, c) \to -\infty$ for $T \to 0$ and $g(T, c) \to \sum_{j=1}^m g_j^R x_j^{a,R}$ from Lemma 4.4 for $T \to \infty$. The first order condition states

$$c = \mu^G(T, R) - T \frac{\partial}{\partial T} \mu^G(T, R) \ .$$

Whether there exists a solution depends on c. Plugging-in (4.10) yields

$$\frac{\partial}{\partial T} \mu^G(T, R) = a' e^{T\Psi^R} g^R \ \Rightarrow \ c = \sum_{j=1}^m g_j^R \int_0^T \left[\mathrm{P}(S_u^R = j) - \mathrm{P}(S_T^R = j) \right] \mathrm{d}u \ .$$

If the right hand side of the upper equation is non-negative for all $T > 0$, the system deteriorates in the sense that states with a high gain rate have a high probability at early times. If the right hand side is non-positive for all $T > 0$, the system rejuvenates and in this case clearly $T = \infty$ is optimal. If we don't know whether there is an 'as-good-as-new' state, we may identify the state which is the preferred initial state. We have to calculate the optimized expected gains per time for every initial state $i = 1, ..., m$. A simultaneous optimization with respect to the initial state i, the replacement time T and the repair matrix R seems again to need much computational effort.

5.2 Age replacement policy

Let $M = (e_1, Q)$ be a failure model with reward rates d. A preventive replacement (the system relaunches in state 1) at costs $c > 0$ is performed at a fixed time $T > 0$ if no failure has occurred before the time to failure τ of the MC J from (3.4). In case of a failure before T the system is replaced at costs $c + k$. A renewal interval has now the length $T \wedge \tau_1$. The expected gains per time unit are

$$g(T, c, k) = \frac{\int_0^T \sum_{j=1}^m d_j \mathrm{P}(J_u = j) \mathrm{d}u - c - k \cdot \mathrm{P}(\tau \le T)}{\int_0^T u f_\tau(u) \mathrm{d}u + T \cdot \mathrm{P}(\tau > T)} \ ,$$

where J is the MC corresponding to τ. The existence of a maximum again depends on the costs c and k. As for the block replacement policy we can check, which working state is closest to a 'as-good-as-new' state, if we also use the other working states as initial state.

5.3 Critical state maintenance policy

Now we present a maintenance policy which takes into account the complex structure of the failure model. We want to identify a set of critical working states of the system and we perform a preventive maintenance action in case the system enters this set. Assume that for every working state j of the system we have a finite set of preventive maintenance actions $\mathcal{P}_j = \{p^{j,1}, ..., p^{j,m_j}\}$ with $p^{j,1} = e_j$, $j = 1, ..., m$. So with $p^{j,1}$ it is always possible that no preventive maintenance is performed. The action $p^{j,i}$ causes costs $\gamma_{ji} \geq 0$, where $\gamma_{j1} = 0$ because nothing is done if we choose the action $p^{j,1}$.

Definition 5.1. Let $(M, \mathcal{R}(M))$ be a repair model with n states and m working states and for $j = 1, ..., m$ let $\mathcal{P}_j = \{p^{i,1}, ..., p^{i,m_j}\} \subset \mathbb{R}^m$ be a non-empty finite set of probability vectors with $p^{j,1} = e_j$. An $m \times m$ matrix P whose rows $(p^{1,l_1})', ..., (p^{m,l_m})'$ are probability vectors $p^{j,l_j} \in \mathcal{R}_j$, $j = 1, ..., m$ is called *preventive maintenance matrix*. We define the set

$$\mathcal{P}(M) := \{P : P \text{ is a preventive maintenance matrix for model } M\}$$

of possible preventive maintenance matrices. We call the triple

$$(M, \mathcal{R}(M), \mathcal{P}(M))$$

a *maintenance model*.

For $j = 1, ..., m$ the row j of a preventive maintenance matrix P defines which preventive maintenance actions is performed when the system enters the state j. We represent a preventive maintenance matrix $P = (p^{1,j_1}, ..., p^{m,j_m})'$ by its index vector $j = (j_1, ..., j_m)$ and use the lexicographical order of these index vectors to order the set $\mathcal{P}(M) = \{P_1, ..., P_{|\mathcal{P}(M)|}\}$.

Definition 5.2. A working state j is called *critical state* under a preventive maintenance matrix $P = (p_{ij})_{i,j=1,...,m}$ if $p_{jj} < 1$, $j = 1, .., m$. The set

$$C(P) := \{j = 1, ..., m | p_{jj} < 1\}$$

is the set of all critical states under P and

$$\overline{C}(P) := \{j = 1, ..., m | p_{jj} = 1\}$$

is the set of all non-critical states under P.

The critical states are the working states at which we perform a preventive maintenance action. Preventive maintenance actions are executed instantaneously. If the system enters a critical state $j \in C(P)$, a multinomial trial with respect to the j-th row of the current preventive maintenance matrix P is accomplished. If the trial results in a critical state $j^* \in C(P)$ again a multinomial trial is accomplished, this time with respect to the j^*-th row of P. This procedure is continued until the state that results from a trial is not critical. Then the system relaunches in this non-critical state.

Definition 5.3. Let $(M, \mathcal{R}(M), \mathcal{P}(M))$ be a maintenance model. A preventive maintenance matrix $P \in \mathcal{P}(M)$ is called *admissible* if being in a critical state the system reaches a non-critical state P-a.s. by the procedure described above.

An admissible preventive maintenance matrix ensures that the system is eventually restarted after it has reached a critical state. Especially if $\overline{C}(P) = \{1, ..., m\}$, a preventive maintenance matrix P is admissible, and if $C(P) = \{1, ..., m\}$ it is not admissible. We want to specify the admissible preventive repair matrices.

Remark 5.1. The case $\overline{C}(P) = \{1, ..., m\}$ (which means that there is no preventive maintenance) is covered by Chapter 4.

Let P be a preventive maintenance matrix with $1 \leq |C(P)| \leq m-1$. By renumbering the states $1, ..., m$ we may write

$$P = \begin{pmatrix} \mathbf{I} & 0 \\ P_1 & P_2 \end{pmatrix} , \tag{5.1}$$

where $\mathbf{I}$ is the identity matrix of dimension $\tilde{m} \times \tilde{m}$ with $\tilde{m} = m - |C(P)| \in \{1, ..., m-1\}$. Now P is admissible if the probability of staying in $C(P)$ tends to zero for increasing number of trials

$$\lim_{k \to \infty} P_2^k = 0 \ .$$

Lemma 5.1. Let P be a preventive maintenance matrix like in (5.1). P is admissible iff $(P_2 - \mathbf{I})$ is regular.

Proof: Let $\lim_{k\to\infty} P_2^k = 0$ and take $x \in \mathbb{R}^{m-\tilde{m}}$ with $(P_2 - \mathbf{I})x = 0$. Now

$$x = P_2 x = P_2^k x \to 0 \ , \text{ for } k \to \infty \ .$$

Thus, $x = 0$ and $(P_2 - \mathbf{I})$ is regular.

Now let $(P_2 - \mathbf{I})$ be regular and let $A \subseteq C(R)$ be an irreducible class of recurrent states. Then the part of P belonging to A is a stochastic matrix $P|_A$ satisfying $P|_A \mathbf{1} = \mathbf{1}$. This means there exists a vector $x \in \mathbb{R}^{m-\tilde{m}}$ with $P_2 x = x$ what contradicts the regularity of $(P_2 - \mathbf{I})$. Hence all states in $C(R)$ are transient and $P_2^k \to 0$ for $k \to \infty$. □

Remark 5.2. The number of preventive maintenance actions performed until the system reaches a non-critical state has a discrete PH-distribution with representation (e_i, P_2) with $e_i \in \mathbb{R}^{m-\tilde{m}}$ (cf. Neuts [45]).

Similarly to the construction in Definition 4.2, applying a repair matrix R and an admissible preventive maintenance matrix P yields a marked point process $(\tau_k^{R,P}, U_k^{R,P})_{k\in\mathbb{N}}$ (the maintenance times $\tau_k^{R,P}$ and corresponding failure types respective critical states $U_k^{R,P}$) and a homogeneous MC that gives the current non-critical state of the system. Let wlg $\overline{C}(P) = \{1, ..., \tilde{m}\}$. The probability of restarting in state $j \in \overline{C}(P)$ after reaching a state $i \in \{\tilde{m}+1, ..., n\}$ is obtained with the matrix

$$\tilde{\Upsilon}(R, P) := \begin{pmatrix} \mathbf{I} & 0 & 0 \\ P_1 & P_2 & 0 \\ R_1 & R_2 & 0 \end{pmatrix} ,$$

where $\mathbf{I}$ is the identity matrix of dimension $\tilde{m} \times \tilde{m}$ and $R = (R_1 \; R_2)$. The sought-after probability is the entry (i, j) of the limiting matrix

$$\tilde{\Upsilon}(R, P)^* := \lim_{k\to\infty} \tilde{\Upsilon}(R, P)^k.$$

Lemma 5.2. The limiting matrix $\tilde{\Upsilon}(R, P)^* := \lim_{k\to\infty} \tilde{\Upsilon}(R, P)^k$ is given by

$$\tilde{\Upsilon}(R, P)^* = \begin{pmatrix} \mathbf{I} & 0 & 0 \\ (\mathbf{I} - P_2)^{-1} P_1 & 0 & 0 \\ R_1 + R_2(\mathbf{I} - P_2)^{-1} P_1 & 0 & 0 \end{pmatrix} .$$

Proof: By induction we obtain for $k \geq 3$

$$\tilde{\Upsilon}(R, P)^k = \begin{pmatrix} \mathbf{I} & 0 & 0 \\ (\mathbf{I} + P_2 + P_2^2 + \cdots + P_2^{k-1}) P_1 & P_2^k & 0 \\ R_1 + R_2(\mathbf{I} + P_2 + P_2^2 + \cdots + P_2^{k-2}) P_1 & R_2 P_2^{k-1} & 0 \end{pmatrix} .$$

Since $(\mathbf{I} - P_2)$ is invertible and

$$\lim_{k\to\infty} P_2^k = 0$$

we have

$$\mathbf{I} + P_2 + P_2^2 + \cdots + P_2^{k-1} = (\mathbf{I} - P_2^k)(\mathbf{I} - P_2)^{-1} \to (\mathbf{I} - P_2)^{-1} \text{ for } k \to \infty \ .$$

□

We denote the essential rows of $\tilde{\Upsilon}(R,P)^*$ as probability vectors $\upsilon^*_{\tilde{m}+1}, ..., \upsilon^*_n$ with

$$\tilde{\Upsilon}(R,P)^* = \left(\begin{array}{c|c} \begin{array}{c} \mathbf{I} \\ \upsilon^*_{\tilde{m}+1} \\ \vdots \\ \upsilon^*_n \end{array} & 0 \end{array} \right)$$

Taking the MC J from Definition 3.1 the initial distribution a may cause J to start in a critical state. This means that already at time $\tau_0 = 0$ a preventive maintenance action is carried out. This special type of maintenance will be considered later, now we only need the real initial distribution a^* after the possible repair action at time zero. The initial distribution a^* on $\{1, ..., \tilde{m}\}$ is given by

$$(a^*)' = a' \left(\begin{array}{c} \mathbf{I} \\ (\mathbf{I} - P_2)^{-1} P_1 \end{array} \right) \ .$$

Definition 5.4. We define a marked point process $(\tau_k^{R,P}, U_k^{R,P})_{k \in \mathbb{N}}$ and a process $S^{R,P} = (S_t^{R,P})_{t \geq 0}$ using a repair matrix $R \in \mathcal{R}(M)$, an admissible preventive maintenance matrix $P \in \mathcal{P}$ with $\overline{C}(P) = \{1, ..., \tilde{m}\}$ $(1 \leq \tilde{m} \leq m-1)$ and copies $\left(J_t^{(k)}\right)_{t \geq 0}$, $k = 1, ...$ of $(J_t)_{t \geq 0}$ from Definition 3.1 with different initial distributions $a^{(k)} \in \mathbb{R}^{\tilde{m}}$ (let $\tau_0^{R,P} = 0$):

- $a^{(1)} := a^*$, $\tau_1^{R,P} := \inf\{t \geq 0 : J_t^{(1)} > \tilde{m}\}$, $U_1^{R,P} := J^{(1)}_{\tau_1^{R,P}}$,
- for $k = 2, ...$ do

$$\begin{aligned} S_t^{R,P} &:= J^{(k-1)}_{t - \tau_{k-2}^{R,P}} && \text{, for } \tau_{k-2}^{R,P} \leq t < \tau_{k-1}^{R,P} \ , \\ a^{(k)} &:= \upsilon^*_{U_{k-1}^{R,P}} && \ , \\ \tau_k^{R,P} &:= \tau_{k-1}^{R,P} + \inf\{t \geq 0 : J_t^{(k)} > \tilde{m}\} && \ , \\ U_k^{R,P} &:= J^{(k)}_{\tau_k^{R,P} - \tau_{k-1}^{R,P}} && \ . \end{aligned}$$

For the next theorem we split the intensity matrix Q into parts for the non-critical states $\{1, ..., \tilde{m}\}$, the critical states $\{\tilde{m}+1, ..., m\}$ and the failure

states $\{m+1, ..., n\}$

$$Q = \begin{pmatrix} Q_{\tilde{m}} & Q^{(2)} & Q^{(3)} \\ & * & \end{pmatrix} , \tag{5.2}$$

where $Q^{(2)} \in \mathbb{R}^{\tilde{m}\times(m-\tilde{m})}$ and $Q^{(3)} \in \mathbb{R}^{\tilde{m}\times(n-m)}$.

Theorem 5.1. The process $S^{R,P} = (S_t^{R,P})_{t\geq 0}$ is a homogeneous MC with state space $\{1, ..., \tilde{m}\}$, initial distribution a^* and intensity matrix $\Psi^{R,P} = \left(\psi_{ij}^{R,P}\right)_{i,j=1,...,\tilde{m}} \in \mathbb{R}^{\tilde{m}\times\tilde{m}}$ with

$$\Psi^{R,P} = Q_{\tilde{m}} + Q^{(2)}(\mathbf{I} - P_2)^{-1}P_1 + Q^{(3)}(R_1 + R_2(\mathbf{I} - P_2)^{-1}P_1) .$$

The distribution of S_t^R is hence given by:

$$\mathrm{P}(S_t^{R,P} = i) = (a^*)' e^{t\Psi^{R,P}} e_i \quad , \ i = 1, ..., \tilde{m} \ , \ t \geq 0 .$$

Proof: This is an immediate consequence of Remark 4.3 with

$$Q_0 = (Q^{(2)}|Q^{(3)}) \text{ and } R = \begin{pmatrix} (\mathbf{I} - P_2)^{-1}P_1 \\ R_1 + R_2(\mathbf{I} - P_2)^{-1}P_1 \end{pmatrix}$$

from (5.2) respectively from Lemma 5.2. □
The processes defined above do not give informations about all maintenance actions performed. The series of maintenance actions needed to reach a non-critical state is at first only important for the costs of maintenance.

Lemma 5.3. Let $\gamma \in \mathbb{R}^{m-\tilde{m}}$ be the vector of respective costs for the preventive maintenance actions given by an admissible preventive maintenance matrix P and let $c \in \mathbb{R}^{n-m}$ be the vector of respective costs for the repair actions given by a repair matrix R. Then the expected costs $\tilde{\gamma}$ and $\tilde{c}$ needed to reach a non-critical state are given by

$$\tilde{\gamma} = (\mathbf{I} - P_2)^{-1}\gamma \quad , \ \tilde{c} = c + R_2\tilde{\gamma} .$$

Proof: We condition on the first maintenance action performed to obtain

$$\tilde{\gamma} = \gamma + P_2\tilde{\gamma} \quad , \ \tilde{c} = c + R_2\tilde{\gamma} .$$

□

Definition 5.5. Let $(M, \mathcal{R}(M), \mathcal{P}(M))$ be a maintenance model, $R \in \mathcal{R}(M)$, $P \in \mathcal{P}(M)$ admissible with $\overline{C}(P) = \{1, ..., \tilde{m}\}$ and $i \in \{\tilde{m}+1, ..., n\}$. The process $N(R, P, i)$ with

$$N_t(R, P, i) := \sum_{k=1}^{\infty} \mathbb{1}_{[0,t]\times\{i\}}(\tau_k^{R,P}, U_k^{R,P}) \ , \ t \geq 0$$

is the canonical counting process for the maintenance actions of type i given R and P.

The counting processes from Definition 5.5 are defined analogous to the counting processes in Definition 4.5. Hence we obtain an analogous SSM-representation as in Lemma 4.1.

Lemma 5.4. Let $(M, \mathcal{R}(M), \mathcal{P}(M))$ be a maintenance model, $R \in \mathcal{R}(M)$, $P \in \mathcal{P}(M)$ admissible with $\overline{C}(P) = \{1, ..., \tilde{m}\}$, and let $\mathfrak{F}$ be the canonical filtration of the counting processes $N(R, P, \tilde{m}+1), ..., N(R, P, n)$. For $i \in \{\tilde{m}+1, ..., n\}$ the $\mathfrak{F}$-intensity of $N(R, P, i)$ is given by

$$\overline{\lambda^i}_t = \sum_{j=1}^{\tilde{m}} q_{ji} \mathbb{1}_{\{j\}}(S_t^{R,P}) \ , \ t \geq 0$$

and the mean value function of $N(R, P, i)$ is

$$\mu(t, R, P, i) := \mathrm{E}(N_t(R, P, i)) = \int_0^t (a^*)' e^{u\Psi^{R,P}} (Q_2 \mid Q_3) e_{i-\tilde{m}} \, \mathrm{d}u \ .$$

Since the costs caused by a failure or by reaching a critical state are random, the cumulated losses up to time $T > 0$ differ from the case with no preventive maintenance (4.4). For $j \in \{\tilde{m}+1, ..., n\}$ let $C^j(R, P)$ be the random costs occurring after reaching state j and $N_T(R, P, j)$ the number of failures respective preventive maintenance actions of type j. Now the cumulated losses up to time $t > 0$ are

$$L(T, R, P) = C_0 + \sum_{j=\tilde{m}+1}^{n} \sum_{i=1}^{N_T(R,P,j)} C_i^j(R, P) \ , \tag{5.3}$$

where $C_1^j(R, P), C_2^j(R, P), ...$ is a series of random variables having the same distribution as $C^j(R, P)$ and C_0 are the costs for maintenance actions at time zero. If a is the initial distribution of the system, $a|_{C(P)}$ is the part corresponding to the critical states. Analogous to Lemma 5.3 the mean of C_0 is

$$\mathrm{E}(C_0) = (a|_{C(P)})'(\mathbf{I} - P_2)^{-1}\gamma = (a|_{C(P)})'\tilde{\gamma} \ .$$

For calculating the expected losses we need a generalized version of Wald's equation.[1]

Lemma 5.5. Let $C = (C_i)_{i\in\mathbb{N}}$ be a series of identically distributed random variables with finite mean. Let N be an $\mathbb{N}_0$-valued random variable with finite mean and for $n \in \mathbb{N}$ let $\mathbb{1}_{[0,n-1]}(N)$ and C_n be independent. Then

$$\mathrm{E}\left(\sum_{i=1}^{N} C_i\right) = \mathrm{E}(N)\mathrm{E}(C_1) \ .$$

Proof: (due to N. Gaffke)
Wlg C_1 is non-negative (else consider the positive and the negative part of C_1). Now due to monotone convergence (e.g. Billingsley [14] Theorem 16.2)

$$\mathrm{E}\left(\sum_{i=1}^{N} C_i\right) = \mathrm{E}\left(\sum_{i=1}^{\infty} C_i \mathbb{1}_{[i,\infty)}(N)\right) = \sum_{i=1}^{\infty} \mathrm{E}(C_i \mathbb{1}_{[i,\infty)}(N)) \ .$$

The random variables C_i and $\mathbb{1}_{[i,\infty)}(N) = 1 - \mathbb{1}_{[0,i-1]}(N)$ are independent, hence

$$\begin{aligned}\mathrm{E}\left(\sum_{i=1}^{N} C_i\right) &= \sum_{i=1}^{\infty} \mathrm{E}(C_i)\mathrm{E}(\mathbb{1}_{[i,\infty)}(N)) \\ &= \mathrm{E}(C_1)\sum_{i=1}^{\infty} \mathrm{P}(N \geq i) \\ &= \mathrm{E}(C_1)\mathrm{E}(N) \ .\end{aligned}$$

□

Concerning (5.3) we notice that the costs $C_i^j(R,P)$ and the process $N(R,P,j)$ need not be independent, but $C_i^j(R,P)$ may only influence the time until one reaches state j the next time. Especially $C_i^j(R,P)$ is independent of the past and therefore independent of $\mathbb{1}_{[0,i-1]}(N_t(R,P,j))$. Applying Lemma 5.5 yields the expected losses up to time T

$$\begin{aligned}\mathrm{E}(L(T,R,P)) = (a|_{C(P)})'\tilde{\gamma} &+ \sum_{j=1}^{m-\tilde{m}} \tilde{\gamma}_j \mu(T,R,P,j+\tilde{m}) \\ &+ \sum_{j=1}^{n-m} \tilde{c}_j \mu(T,R,P,j+m) \ .\end{aligned}$$

[1] Wald's equation (cf. Blackwell [15]) gives the result of Lemma 5.5 with the additional assumption, that N is a stopping time with respect to the canonical filtration of the iid series C.

The expected income up to time T is similar to (4.10)

$$\mu^I(T,R,P) = \int_0^T \sum_{j=1}^{\tilde{m}} d_j \mathrm{P}(S_u^{R,P} = j)\ \mathrm{d}u\ .$$

Altogether we calculate the expected gains

$$\begin{aligned}\mu^G(T,R,P) = &-(a|_{C(P)})'\tilde{\gamma} \\ &+ \sum_{i=1}^{\tilde{m}} \int_0^T \mathrm{P}(S_u^{R,P} = i)\left[d_i - (e_i)'Q^{(2)}\tilde{\gamma} - (e_i)'Q^{(3)}\tilde{c}\right]\ \mathrm{d}u\ .\end{aligned} \tag{5.4}$$

Remark 5.3. Since it always holds that $\tilde{c} \geq c$ (Lemma 5.3), the term

$$g_i^{R,P} := \left[d_i - (e_i)'Q^{(2)}\tilde{\gamma} - (e_i)'Q^{(3)}\tilde{c}\right]\ ,\ i = 1, ..., \tilde{m}$$

from (5.4) is never higher than the gain rate g_i^R from (4.11) in the model without preventive maintenance. Hence a preventive maintenance action only yields higher expected gains than without preventive maintenance, if the probability that the system is in a state with a high gain rate is higher than before. $g_i^{R,P}$ is the gain rate for the maintenance model.

Definition 5.6. Let $(M, \mathcal{R}(M), \mathcal{P}(M)$ be a maintenance model. We call a repair matrix $R^* \in \mathcal{R}(M)$ and an admissible preventive maintenance matrix $P^* \in \mathcal{P}(M)$

i) optimal up to time $T > 0$, if

$$\mu^G(T,R^*,P^*) \geq \mu^G(T,R,P)\ \forall\ R \in \mathcal{R}(M) \text{ and admissible } P \in \mathcal{P}(M)\ .$$

ii) finally optimal, if

$$\begin{aligned}\lim_{T\to\infty} T^{-1}\mu^G(T,R^*,P^*) \geq \lim_{T\to\infty} T^{-1}\mu^G(T,R,P)\\ \forall\ R \in \mathcal{R}(M) \text{ and admissible } P \in \mathcal{P}(M)\ .\end{aligned}$$

We call

$$\Upsilon(R,P) := \begin{pmatrix} \mathbf{I} & 0 \\ P_1 & P_2 \\ R_1 & R_2 \end{pmatrix}$$

a *maintenance matrix*.

The index vector $j = (i_1, ..., i_m, l_1, ..., l_{n-m})'$ of a maintenance matrix $\Upsilon(R, P)$ is given by the index vector $i = (i_1, ..., i_m)'$ of $P = (p^{1,i_1}, ..., p^{m,i_m})'$ and the index vector $l = (l_1, ..., l_{n-m})'$ of $R = (r^{1,l_1}, ..., r^{n-m,l_{n-m}})'$. We order the possible maintenance matrices by the lexicographical order of the respective index vectors. We notice that the structure of the expected gain function (5.4) is the same as in 4.10. This is why we may use the first three approaches from Chapter 4 in order to obtain an optimal maintenance policy (R, P). We illustrate these approaches with an example.

Example 5.2. We consider the repair models from Example 4.6.

i) For the repair model from Example 4.6 we assume the possible preventive maintenance actions

$$\mathcal{P}_1 = \{e_1, e_2\} \ , \ \mathcal{P}_2 = \{e_2, e_1\} \ , \ \mathcal{P}_3 = \left\{ e_3, \begin{pmatrix} 0.5 \\ 0.5 \\ 0 \end{pmatrix} \right\} \ ,$$

and the corresponding costs

$$\Gamma = \begin{pmatrix} 0 & 0.1 \\ 0 & 0.3 \\ 0 & 0.5 \end{pmatrix} \ .$$

There are 8 possible replacement matrices, but the matrices

$$P_7 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \ , \ P_8 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0.5 & 0.5 & 0 \end{pmatrix}$$

are not admissible. The expected gains (5.4) for $T = 1$ and $T = 25$ are given in Figure 5.2.

In both cases the maintenance matrices $\Upsilon_{49}, ..., \Upsilon_{64}$ are optimal. For all these matrices the only non-critical state is state 1. The only possible failure state is state 4, so if we change the rows $2, ..., 5$ of the repair matrix R, it does not affect the expected gains. The maintenance matrices $\Upsilon_{49}, ..., \Upsilon_{64}$ are also finally optimal, and since $\mathrm{P}(S_t^{R,P} = 1) = 1 \ , \ t \geq 0$, the expected gains (5.4) for these maintenance matrices equal

$$\mu^G(T, R, P) = 97.45T \ , \ T \geq 0 \ .$$

We simulate $k = 100$ cumulated gains, and similar to Example 4.9 we calculate confidence intervals for the expected gains. The results are

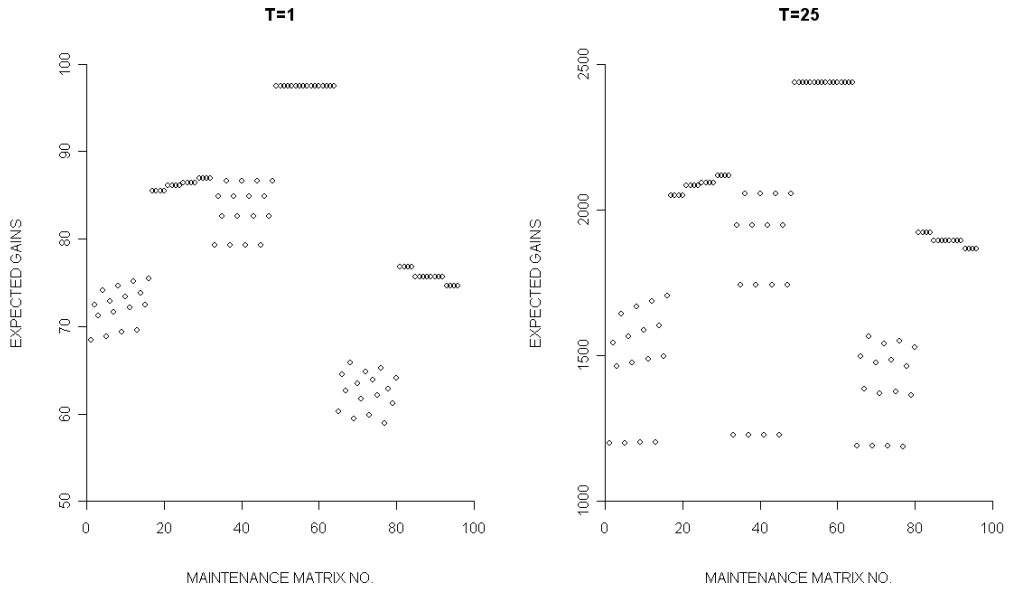


Figure 5.2: Expected gains for the maintenance model in Example 5.2 i)

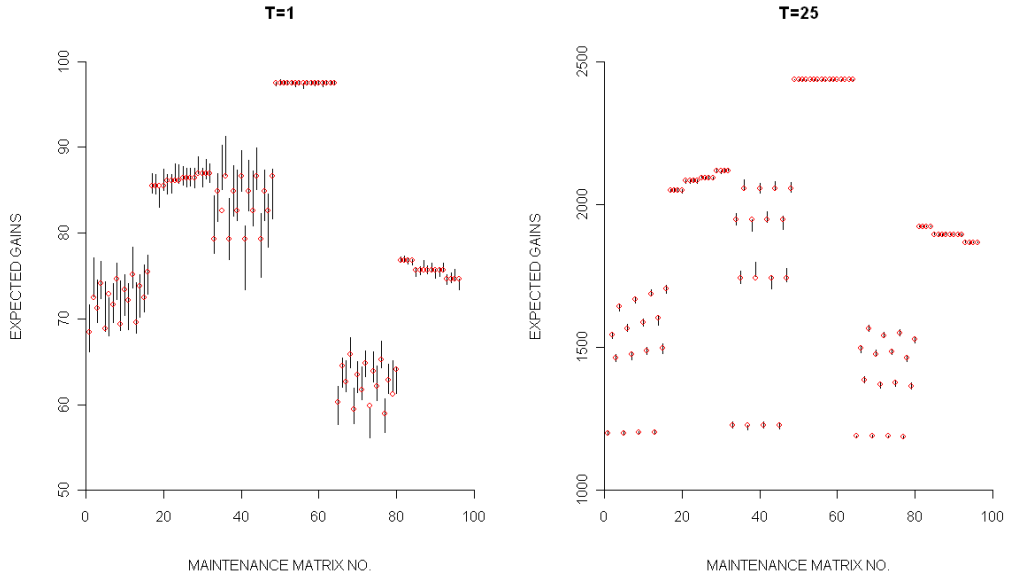


Figure 5.3: Confidence intervals for the expected gains, Example 5.2 i)

shown in Figure 5.3, where the red dots are the respective true parameters.

ii) For the repair model in Example 4.6 ii) we assume that the possible preventive maintenance actions are all *determined preventive maintenance actions*:

$$\mathcal{P}_i = \{e_1, ..., e_m\} \ , \ i = 1, ..., m \ .$$

Note that due to computational reasons we differ slightly from Definition 5.1, since we have $p^{i,i} = e_i$ instead of $p^{i,1} = e_i$ in the definition. The corresponding costs are given by

$$\Gamma = \begin{pmatrix} 0 & 0.1 & 0.1 \\ 0.1 & 0 & 0.1 \\ 0.1 & 0.1 & 0 \end{pmatrix} \ .$$

The matrix $\Gamma = (\gamma_{ij})_{i,j=1,...,m}$ in turn coincides with Definition 5.1, because for $i = 1, ..., m$ the preventive maintenance action $e_i \in \mathcal{P}_i$ causes costs $\gamma_{ii} = 0$. There are 27 possible preventive maintenance matrices, but 11 of them are not admissible, namely the $2^3 = 8$ matrices P with $C(P) = \{1, 2, 3\}$ (these are $P_{10}, P_{11}, P_{16}, P_{17}, P_{19}, P_{20}, P_{25}, P_{26}$) and the 3 matrices with $C(P) = \{j_1, j_2\}$ and $i_{j_1} = j_2$ and $i_{j_2} = j_1$, where $i = (i_1, i_2, i_3)'$ is the index vector of P (these are the matrices P_8, P_{12}, P_{22}). The expected gains (5.4) are for $T = 1$ and $T = 25$ are given in Figure 5.4. In both cases the maintenance matrices $\Upsilon_1, ..., \Upsilon_{81}$ are optimal. Again, for all these matrices the only non-critical state is state 1 and the only possible failure state is state 4. The maintenance matrices $\Upsilon_1, ..., \Upsilon_{81}$ are also finally optimal, and since $\mathrm{P}(S_t^{R,P} = 1) = 1 \ , \ t \geq 0$, the expected gains (5.4) for these maintenance matrices equal

$$\mu^G(T, R, P) = -0.4T \ , \ T \geq 0 \ .$$

We simulate $k = 100$ cumulated gains, and similar to Example 4.9 we estimate the expected gains. The results are shown in Figure 5.5.

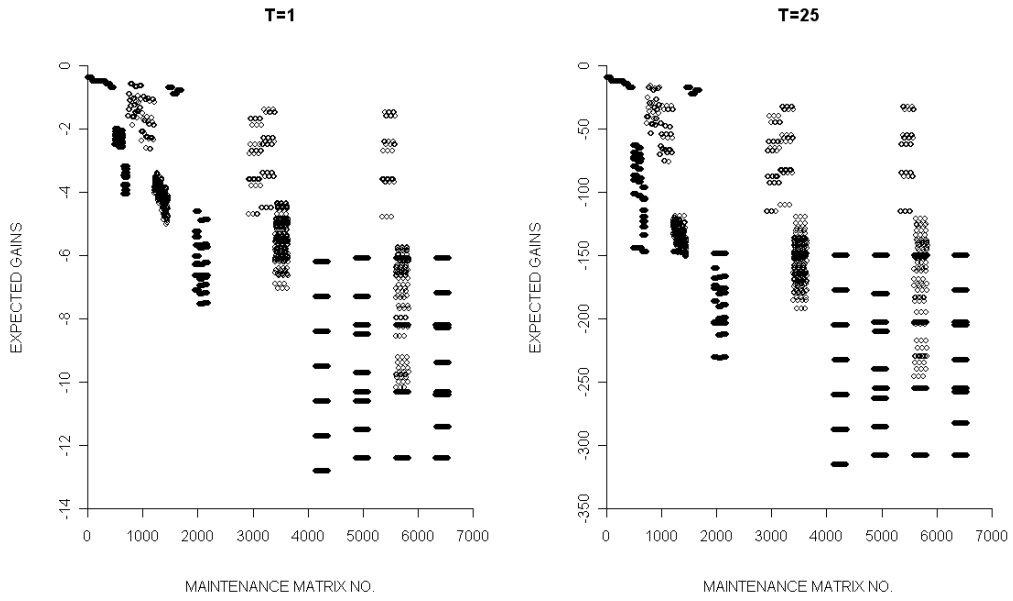


Figure 5.4: Expected gains for the maintenance model in Example 5.2 ii)

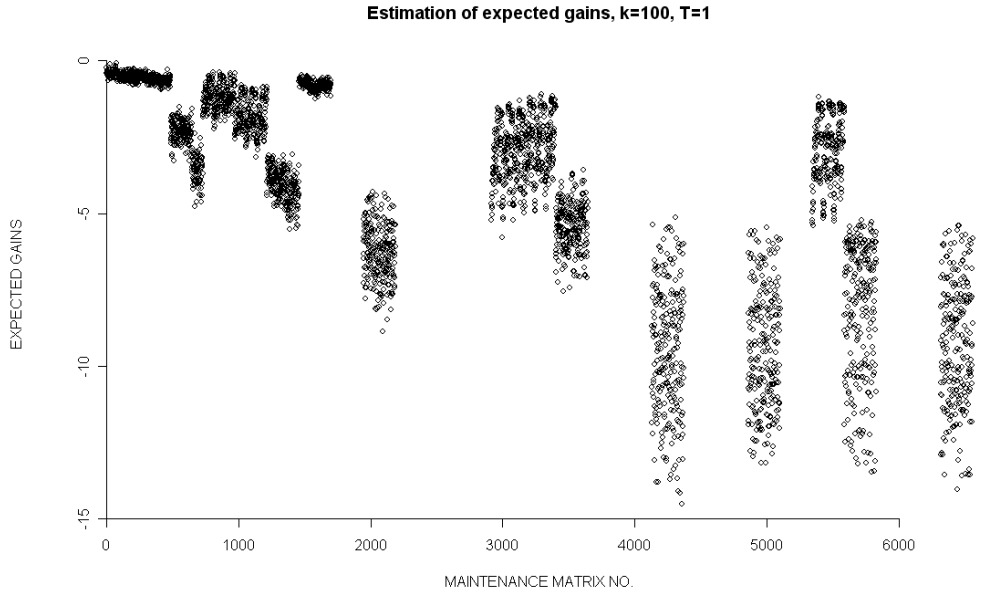


Figure 5.5: Estimation of the expected gains, Example 5.2 ii)

We may draw some conclusions from this example.

- There are maintenance matrices that yield higher expected gains than without preventive maintenance, but there are also maintenance matrices that yield lower expected gains (cf. Chapter 4).
- There are preventive maintenance matrices which are not admissible. At least there are $(m_1 - 1) \cdot ... \cdot (m_m - 1)$ matrices P with $C(P) = \{1, ..., m\}$ (where $m_j = |\mathcal{P}_j|$, $j = 1, ..., m$).
- If two maintenance matrices only differ at rows, which correspond to states that never occur, they yield the same expected gains.

We may again avoid to calculate the expected costs for all possible maintenance matrices, if we use a sequential approach. We present and afterwards discuss an algorithm, which does not take into account multiple simultaneous maintenance actions. Let $D = \text{diag}(d_1, ..., d_m)$, $V = \text{diag}(-q_{11}^{-1}, ..., -q_{mm}^{-1})$, $v = (-q_{11}^{-1}, ..., -q_{mm}^{-1})'$ and $W = (VQ_m + \mathbf{I}, VQ_0)$. So v is the vector of waiting times until the next jump and W is basically the one-step transition matrix (2.10) of the MC J. We use the following algorithm to optimize the repair matrix and the preventive maintenance policy simultaneously:

(Step 1) for each failure type $j = m + 1, ..., n$ do:

$$\begin{aligned} l^* &:= \underset{l=1,...,n_{j-m}}{\text{argmax}} \frac{(r^{j-m,l})' Dv - c_{j-m,l}}{(r^{j-m,l})' v} \ , \\ r_{j-m}^{(1)} &:= l^* \ , \\ G_{j-m}^{(1)} &:= (r^{j-m,l^*})' Dv - c_{j-m,l^*} \ , \\ T_{j-m}^{(1)} &:= (r^{j-m,l^*})' v \ . \end{aligned}$$

for each running state of the system $i = 1, ..., m$ do:

$$\begin{aligned} l^* &:= \underset{l=1,...,m_i}{\text{argmax}} \frac{(p^{j,l})' Dv - \gamma_{i,l}}{(p^{i,l})' v} \ , \\ p_i^{(1)} &:= l^* \ , \\ \Gamma_i^{(1)} &:= (p^{j,l^*})' Dv - \gamma_{i,l^*} \ , \\ \Theta_i^{(1)} &:= (p^{i,l^*})' v \ . \end{aligned}$$

(Step s) for each failure type $j = m+1, ..., n$ do:

$$l^* := \underset{l=1,...,n_{j-m}}{\operatorname{argmax}} \frac{(r^{j-m,l})'Dv - c_{j-m,l} + (r^{j-m,l})'W \begin{pmatrix} \Gamma^{(s-1)} \\ G^{(s-1)} \end{pmatrix}}{(r^{j-m,l})'v + (r^{j-m,l})'W \begin{pmatrix} \Theta^{(s-1)} \\ T^{(s-1)} \end{pmatrix}} ,$$

$$r_{j-m}^{(s)} := l^* ,$$

$$G_{j-m}^{(s)} := (r^{j-m,l^*})'Dv - c_{j-m,l^*} + (r^{j-m,l^*})'W \begin{pmatrix} \Gamma^{(s-1)} \\ G^{(s-1)} \end{pmatrix} ,$$

$$T_{j-m}^{(s)} := (r^{j-m,l^*})'v + (r^{j-m,l^*})'W \begin{pmatrix} \Theta^{(s-1)} \\ T^{(s-1)} \end{pmatrix} .$$

for each running state of the system $i = 1, ..., m$ do:

$$l^* := \underset{l=1,...,m_i}{\operatorname{argmax}} \frac{(p^{j,l})'Dv - \gamma_{i,l} + (p^{i,l})'W \begin{pmatrix} \Gamma^{(s-1)} \\ G^{(s-1)} \end{pmatrix}}{(p^{i,l})'v + (p^{i,l})'W \begin{pmatrix} \Theta^{(s-1)} \\ T^{(s-1)} \end{pmatrix}} ,$$

$$p_i^{(s)} := l^* ,$$

$$\Gamma_i^{(s)} := (p^{j,l^*})'Dv - \gamma_{i,l^*} + (p^{i,l^*})'W \begin{pmatrix} \Gamma^{(s-1)} \\ G^{(s-1)} \end{pmatrix} ,$$

$$\Theta_i^{(s)} := (p^{i,l^*})'v + (p^{i,l^*})'W \begin{pmatrix} \Theta^{(s-1)} \\ T^{(s-1)} \end{pmatrix} .$$

(STOP) Stop if for a fixed integer $K \geq 1$:

$$\exists\ s_1, ..., s_K :\ 1 \leq s_1 < ... < s_K\ ,\ r^{(s_1)} = ... = r^{(s_K)}\ ,\ p^{(s_1)} = ... = p^{(s_K)}\ .$$

Then choose $R = (r^{1,j_1}|...|r^{n-m,j_{n-m}})'$ with $j_i = r_i^{(s_1)}$, $i = 1, ..., n-m$ as repair matrix, and choose $P = (p^{1,j_1}|...|p^{n-m,j_{n-m}})'$ with $j_i = p_i^{(s_1)}$, $i = 1, ..., m$ as preventive maintenance matrix. If the system enters state $i \in C(P)$ we perform the preventive maintenance action p^{i,j_i}..

This algorithm yields a maintenance policy which was optimal in K steps. We will now use the algorithm for some examples.

Example 5.3. We consider the maintenance models from Example 5.2 with $K = 5$.

i) The steps of the algorithm yield the following maintenance matrices:

Step	j_1	j_2	j_3	j_4	j_5	j_6	j_7	j_8	no.
1	1	2	2	1	2	2	2	2	64
2	1	2	2	1	2	2	2	2	64
3	1	2	2	1	2	2	2	2	64
4	1	2	2	1	2	2	2	2	64
5	1	2	2	1	2	2	2	2	64

The maintenance matrix Υ_{64} is obtained in all steps of the algorithm. According to Example 5.2, the maintenance matrix Υ_{64} is finally optimal and optimal in $[0, 1]$ and $[0, 25]$.

ii) The steps of the algorithm yield the following repair matrices:

Step	j_1	j_2	j_3	j_4	j_5	j_6	j_7	j_8	no.
1	1	2	3	3	3	3	3	3	1458
2	1	1	1	1	2	2	2	2	41
3	1	1	1	1	2	2	2	2	41
4	1	1	1	1	2	2	2	2	41
5	1	1	1	1	2	2	2	2	41
6	1	1	1	1	2	2	2	2	41

The resulting maintenance matrix Υ_{41} is known to be finally optimal and also optimal in $[0, 1]$ (Example 5.2). The maintenance matrix Υ_{1458}, which is obtained in the first step of the algorithm, equals the repair matrix R_{243} in Example 4.7.

For Example 5.3 we know, that the algorithm works well, because we may compare the results with the exact calculations. Now we consider an example which seems to be not manageable for the first three approaches, because the number of possible maintenance matrices is large.

Example 5.4. We consider the determined maintenance model for a system with $m = 10$ working states and $n - m = 19$ failure states. The number of possible maintenance matrices equals 10^{29}. We assume the parameters

$a = e_1$, $d = \mathbf{1}$ and

$$Q_{10} = \begin{pmatrix} -3 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -5 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & -5 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -2 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -6 & 0 & 0 & 0 & 2 \\ 1 & 0 & 1 & 0 & 0 & 0 & -5 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & -10 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & -8 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -2 \end{pmatrix},$$

$$Q_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

We consider equal costs for repairs $C = (c_{ij})_{i,j=1,...,10}$ with $c_{ij} = 1$, $i, j = 1, ..., 10$. The crucial point of the algorithm is, that the critical states are not incorporated. Therefore we consider two different costs matrices for the preventive maintenance actions $\Gamma^{(1)}$ and $\Gamma^{(2)}$.

$$\Gamma^{(1)} = \begin{pmatrix} 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 0.0 \end{pmatrix},$$

$$\Gamma^{(2)} = \begin{pmatrix} 0.0 & 0.01 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.0 & 0.01 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.0 & 0.01 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.0 & 0.01 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.0 & 0.01 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.0 & 0.01 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.0 & 0.01 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.0 & 0.01 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.0 & 0.01 \\ 0.01 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.0 \end{pmatrix}.$$

The effect of the costs for repairs is that at one step the algorithm yields the same repair action for all failure types. Whereas for $\Gamma^{(1)}$ all preventive maintenance actions are equal, the costs in $\Gamma^{(2)}$ show a path how to reach a preferable state. We display the results from the algorithm with $K = 10$, and we only show steps, at which the maintenance matrix changes, together with the corresponding final behavior $(x^{a,R,P})'g^{R,P}$ of the expected gains (5.4). For $\Gamma^{(1)}$ we obtain

Step	j_1	j_2	j_3	j_4	j_5	j_6	j_7	j_8	j_9	j_{10}	j_{11}	$(x^{a,R,P})'g^{R,P}$
1	1	2	3	4	5	6	7	8	9	10	4	-0.495
2	1	1	1	4	1	1	1	1	1	10	4	-0.15
4	1	4	4	4	4	4	4	4	4	10	4	-0.1

.

The results make perfect sense. Since all preventive maintenance actions cost the same, we jump to the most preferable state (in this case state 4). Only if a state is already close to the most preferable state, we perform no preventive maintenance action (in this case for the states 1 and 10). On the other hand we obtain for $\Gamma^{(2)}$

Step	j_1	j_2	j_3	j_4	j_5	j_6	j_7	j_8	j_9	j_{10}	j_{11}	$(x^{a,R,P})'g^{R,P}$
1	1	2	3	4	5	6	7	8	9	10	4	-0.495
2	1	3	4	4	6	7	7	9	10	1	4	-0.03
3	1	3	4	4	1	7	7	1	10	1	4	-0.233
11	1	3	4	4	1	7	7	1	10	1	1	-0.165

.

These results are nonsense. First, the expected gains for the resulting maintenance matrix are lower than the expected gains for the maintenance matrix obtained in step 2. Second, and this is the structural problem, it makes absolutely no sense to e.g. go directly to state 1 if one enters state 8 at costs $\gamma_{81} = 0.3$. It would be cheaper to perform multiple preventive maintenance actions, first go to state 9, then to state 10, then to state 1 at costs $\gamma_{89} + \gamma_{9,10} + \gamma_{10,1} = 0.03$.

Summing up, we may say that there is evidence, that the algorithm results in a suitable maintenance matrix, if the model is reasonable. In the determined maintenance case, a reasonable model should always fulfill $\gamma_{ij} \leq \gamma_{ik} + \gamma_{kj}$, $i, j, k = 1, ..., m$. There is no guarantee, that the resulting preventive maintenance matrix is admissible. As a help to hand, one should always compare the expected gains of the resulting maintenance matrix with the expected gains corresponding to the other maintenance matrices, which appear at the steps of the algorithm.

Chapter 6

Conclusions

We summarize the properties of the maintenance model presented in the Chapters 3 to 5.

Chapter 3: The failure model

- We consider PH-distributed times to failure. The class of PH-distributions is rich enough to ensure a broad field of applications. Nevertheless, the PH-distributions are mathematical tractable by using matrix calculus.
- Since the working states of the system are described by the condition of the system, the states allow a straightforward verbal interpretation.
- We allow different failure types. Again, they may be described verbally.
- In many applications one observes only a finite number of different conditions of the system. The condition of an engine is given by warning lights, and the condition of a human being is given by a medical diagnosis.
- The assumption that changes of the system's condition occur due to random shocks and not continuously may be plausible in some applications.

Chapter 4: The repair model

- The repair model may offer a variety of different possible repair actions, including repair actions of a different quality.
- The repair policies minimal repair, replacement and imperfect repair are included (if these are possible repair actions).

- We offer a flexible optimality criterion. State-dependent rewards are typical characteristics of systems subject to ageing.

- Exact calculations are possible, but extensive.

- We offer a useful heuristic approach for finding an optimal repair policy.

Chapter 5: The maintenance model

- The critical state replacement policy seems to be more feasible for this model than the block and the age replacement policy.

- There is again a straightforward interpretation: If the system enters a critical state, one should immediately start with a countermeasure. Therefore, the maintenance policy is easy to apply.

- We simultaneously optimize the repair policy and the preventive maintenance policy.

- Exact calculations are possible, but even more extensive.

- We offer a useful heuristic approach for reasonable models, but the general behavior is unacquainted.

Potential topics for further research are

- If we consider hierarchical structures like acyclic failure models, are there options to improve the presented methods?

- There are many approaches for the calculation of matrix exponentials. Especially for sparse matrices, are there better methods than the one presented?

- How may we use concepts from risk theory (e.g. ruin probabilities for the cumulated gains)?

- How can one modify the heuristic approaches, if one has only partial information?

- Are there better algorithms for choosing a maintenance policy?

- There is a broad theory for renewal processes. How can one apply these techniques for our model?

- How may we estimate the parameters of the model?

Concerning the last point we give a brief review. Assume that we have observed the processes $(\tau_n^R, U_n^R)_{n\in\mathbb{N}}$ and $(S_t^R)_{t\geq 0}$ from Definition 4.2 in the interval $[0,T]$. How should we estimate Q?

The first step is to estimate the generator Ψ^R of $(S_t^R)_{t\geq 0}$. Maximum-likelihood estimation of the generator of a Markov chain has been examined by e.g. Billingsley [13] and Jacobsen [29]. One condition on the regularity of the likelihood function is, that Ψ^R is irreducible which is not fulfilled in this model. Anyway, the MLE under regularity gives us an idea how to estimate in this model.
Let $(T_1, ..., T_N)$ resp. $(\Delta_1, ..., \Delta_N)$ be the jump times respectively the inter-jump times of S^R in $[0,T]$, Δ_{N+1} the next inter-jump time and $Y_0 = S_0^R$, $Y_n = S_{T_n}^R$ for $n = 1, ..., N$. The likelihood function given one path $\{S_t^R : 0 \leq t \leq T\}$ is

$$L(\Psi^R) = \prod_{i=1}^{m} \prod_{j\neq i} (\Psi_{ij}^R)^{N_{ij}(T)} \exp\left(-\Psi_{ij}^R \int_0^T \mathbb{1}_{\{i\}}(S_t^R)\mathrm{d}t\right),$$

where $N_{ij}(T)$ is the number of jumps from state i to state j in $[0,T]$.

The maximum likelihood estimator of $\Psi^R = (\Psi_{ij}^R)_{i,j=1,...,m}$ is (when defined):

$$\widehat{\Psi^R}_{ij} = \frac{N_{ij}(T)}{\int_0^T \mathbb{1}_{\{i\}}(S_t^R)\mathrm{d}t}\ , \quad i \neq j \text{ and } \widehat{\Psi^R}_{ii} = -\sum_{j\neq i} \widehat{\Psi^R}_{ij}\ .$$

Let now $\tilde{N}_{ij}(T)$ be the number of failures τ_k^R of type $U_k^R = j$ with $S_{\tau_k^R -}^R = i$ in $[0,T]$. We plug-in and obtain an estimator for Q_0:

$$\widehat{Q}_{ij} = \frac{\tilde{N}_{ij}(T)}{\int_0^T \mathbb{1}_{\{i\}}(S_t^R)\mathrm{d}t}\ , \quad i = 1, ..., m,\ j = m+1, ..., n\ .$$

Denote this matrix by $\widehat{Q}_0$ and plug-in for an estimator for Q_m:

$$\widehat{Q}_m = \widehat{\Psi^R} - \widehat{Q}_0 \cdot R\ .$$

This should give a first impression about statistical inference for this model and we advert to some different approaches given by Asmussen et al. [7] (EM algorithm), Bladt and Sörensen [16] (discrete observations) and Hjort and Varin [28] (partial likelihood and quasi likelihood).

Bibliography

[1] Aalen, O. *Statistical Inference for a Family of Counting Processes.* PhD thesis, University of California, Berkeley, 1975.

[2] Aalen, O. Phase type distributions in survival analysis. *Scandinavian Journal of Statistics 22* (1995).

[3] Aalen, O., and Gjessing, H. Understanding the shape of the hazard rate: A process point of view. *Statistical Science, Vol.16, No.1* (2001).

[4] Agaev, R., and Chebotarev, P. On determining the eigenprojection and components of a matrix. *Automation and Remote Control, Vol. 63, No. 10* (2002).

[5] Ahn, S., and Ramaswami, V. Bilateral phase type distributions. *Stochastic Models, Vol. 21, Issue 2,3* (2005).

[6] Asmussen, S. *Ruin Probabilities.* World Scientific, 2000.

[7] Asmussen, S., Nerman, O., and Olsson, M. Fitting phase-type distributions via the em algorithm. *Scandinavian Journal of Statistics 23* (1996).

[8] Aven, T., and Jensen, U. *Stochastic Models in Reliability.* Springer, 1999.

[9] Barlow, R., and Hunter, L. Optimum preventive maintenance policies. *Operations Research, Vol.8, No.1* (1960).

[10] Beichelt, F. A unifying treatment of replacement policies with minimal repair. *Naval Research Logistics, Vol.40* (1993).

[11] Beichelt, F. *Zuverlässigkeits- und Instandhaltungstheorie.* B.G. Teubner Stuttgart, 1993.

[12] Beichelt, F. *Stochastic Processes in Science, Engineering and Finance*. Chapman and Hall, 2006.

[13] Billingsley, P. *Statistical Inference for Markov Processes*. The University of Chicago Press, 1961.

[14] Billingsley, P. *Probability and Measure, 3rd Edition*. John Wiley and Sons, 1995.

[15] Blackwell, D. On an equation of wald. *Annals of Mathematical Statistics, Vol.17, No.1* (1946).

[16] Bladt, M., and Sörensen, M. Statistical inference for discretely observed markov jump processes. *Journal of the Royal Statistical Society, Vol. 67, Part 3* (2005).

[17] Block, H. W., Li, Y., Savits, T. H., and Wang, J. Continuous mixtures with bathtub-shaped failure rates. *Journal of Applied Probability 45* (2008).

[18] Bobbio, A., Cumani, A., Premoli, A., and Saracco, O. Modeling and identification of non-exponential distributions by homogeneous markov processes. In *6th Advances in Reliability Technology Symposium NCSR* (1980).

[19] Bosch, K., and Jensen, U. Instandhaltungsmodelle - eine übersicht. *OR Spektrum 5* (1983).

[20] Brown, M., and Proschan, F. Imperfect repair. *Journal of Applied Probability 20* (1983).

[21] Cox, D. R. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society B, 17* (1955).

[22] Devroye, L. *Non-Uniform Random Variate Generation*. Springer New York, 1986.

[23] Doob, J. L. *Stochastic Processes*. Wiley, 1953.

[24] Doorn, E., and Pollett, P. Survival in a quasi-death process. *Linear Algebra and its applications 429* (2008).

[25] Dunford, N., and Schwartz, J. *Linear Operators Part I: General Theory*. Wiley Classics Library Edition, 1988.

[26] Finkelstein, M. *Failure Rate Modeling for Reliability and Risk.* Springer London, 2008.

[27] Harris Jr., W. A., Fillmore, J. P., and Smith, D. R. Matrix exponentials - another approach. *SIAM Review, Vol. 43, No. 4* (2001).

[28] Hjort, N., and Varin, C. Ml, pl, ql in markov models. *Scandinavian Journal of Statistics 35* (2008).

[29] Jacobsen, M. *Statistical Analysis of Counting Processes.* Springer New York, 1982.

[30] Jacobsen, M. Martingales and the distribution of the time to ruin. *Stochastic Processes and their Applications, Volume 107, Number 1* (2003).

[31] Jacobsen, M. *Point Process Theory and Applications.* Birkhäuser, 2006.

[32] Jensen, U. A general replacement model. *ZOR - Methods and Models of Operations Research 34* (1990).

[33] Kallenberg, O. *Foundations of Modern Probability, 2nd Edition.* Springer New York, 2002.

[34] Kijima, M. *Markov Processes for Stochastic Modeling.* Chapman and Hall, 1997.

[35] Kijima, M. Hazard rate and reversed hazard rate monotonicities in continuous-time markov chains. *Journal of Applied Probability 35* (1998).

[36] Kronmal, R. A., and Peterson, A. V. The alias and alias-rejection-mixture methods for generating random variables from probability distributions. In *Proceedings of the 11th conference on Winter simulation - Volume 1* (1979).

[37] Last, G., and Brandt, A. *Marked Point Processes on the Real Line.* Springer New York, 1995.

[38] Limnios, N., and Oprisan, G. *Semi-Markov Processes and Reliability.* Birkhäuser, 2001.

[39] Luther, U., and Rost, K. Matrix exponentials and inversion of confluent vandermonde matrices. *ETNA, Volume 18, pp.91-100* (2004).

[40] Marshall, A. W., and Olkin, I. *Life distributions.* Springer New York, 2007.

[41] Meyer, P.-A. A decomposition theorem for supermartingales. *Illinois Journal of Mathematics 6* (1962).

[42] Meyer, P.-A. Decomposition of supermartingales: the uniqueness theorem. *Illinois Journal of Mathematics 7* (1963).

[43] Moler, C., and van Loan, C. Nineteen dubious ways to compute the exponential of a matrix. *SIAM Review, Vol.20* (1978).

[44] Moler, C., and van Loan, C. Nineteen dubios ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review, Vol.45, No.1* (2003).

[45] Neuts, M. F. Probability distributions of phase type. *Liber Amicorum prof. Emeritus H. Florin* (1975).

[46] Neuts, M. F. Renewal processes of phase type. *Naval Research Logistics Quarterly, Vol.25, Issue 3* (1978).

[47] Neuts, M. F. *Matrix-geometric solutions in Stochastic Models.* Dover Publications, 1981.

[48] Neuts, M. F., and Meier, K. S. On the use of phase-type distributions in reliability modelling of systems with two components. *OR Spektrum 2* (1980).

[49] Neuts, M. F., and Pagano, M. E. Generating random variates from a distribution of phase type. In *1981 Winter Simulation Conference Proceedings* (1981).

[50] Norris, J. R. *Markov chains.* Cambridge University Press, 1997.

[51] O'Cinneide, C. A. Characterization of phase-type distributions. *Stochastic Models* (1990).

[52] O'Cinneide, C. A. Phase-type distributions: Open problems and a few properties. *Stochastic Models* (1999).

[53] Serfozo, R. *Probability and its applications.* Springer, 2009.

[54] Turner, R. L. Inverse of the vandermonde matrix with applications. *NASA Technical Note D-3547* (1966).

[55] Walker, A. J. An efficient method for generating discrete random variables with general distributions. *ACM Transactions on Mathematical Software, Volume 3, Issue 3* (1977).