Methods for Approximating the Availability Functions

Version of July 25, 2008

Łukasz Bednara

Methods for Approximating the Availability Functions

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

APPLIED MATHEMATICS

by

Łukasz Bednara
born in Lublin, Poland

Risk and Environmental Modeling,
Department of Applied Mathematics,
Delft University of Technology
Delft, the Netherlands
www.ewi.tudelft.nl

Shell Global Solutions International
PO Box 38000, 1030 BN
Amsterdam, the Netherlands
www.shell.com/globalsolutions/

Methods for Approximating the Availability Functions

Author: Łukasz Bednara
Student id: 1320696
Email: LukaszBednara@gmail.com

Abstract

Availability studies aim at modeling and quantifying the relation between system design and production effectiveness. Once this relation is modeled, different designs and operating strategies can be compared and ranked. Usually, the objective is to estimate and minimize losses that are due to equipment failures and planned shutdowns. Future performance of each unit/component may be modeled by the availability function that gives the probability that the unit is functioning at time t. Those functions are then combined together to assess future availability of a plant.

The main aim of this project was to improve existing methods for computing the availability functions on a component level. Formulation of the problem in a general framework allowed to obtain extra availability metrics that include the present condition of a component. Different methods for computing availability functions were tested and compared with respect to speed and accuracy. The best method was selected based on the performance on different problem families. For the purpose of the comparison the explicit formulas for the availability functions in case of Gamma alternating renewal process were derived. Additionally some new bounds on the availability functions were obtained. The last part presents a new approach for modeling the availability of a component in situation when some failures are not visible (so called ”grace periods”).

Thesis Committee:

Responsible professor:	Prof. Michel Dekking, Faculty EEMCS, TU Delft
University supervisor:	Dr. J.A.M van der Weide, Faculty EEMCS, TU Delft
Company supervisor:	Dr. ir. Martina Kramer, Shell Global Solutions
Committee Member:	Prof. Rommert Dekker, Faculty of Economics, Erasmus University

Preface

This project was done for the Availability Team at Shell Global Solutions International as a partial fulfillment of the requirements for the degree Master of Science. I worked on this project for half of my two-year stay in Delft. This was unforgettable time and I would like to show my gratitude to the people that were together with me during this period.

I am very grateful to prof. Jolanta Misiewicz, prof. Roger M. Cooke and Dr Dorota Kurowicka, for giving me this fantastic chance to follow the Risk and Environmental Modeling program at TU Delft.

I would like to thank to Dr Hans van der Weide, prof. Michel Dekking and prof. Rommert Dekker for help, inspiration, corrections and comments. I am also grateful to prof. Michael Tortorella for his interest in the project.

I thank people from Shell Global Solutions for giving me the opportunity to work on this interesting and challenging project. Especially, I am grateful to Martina Kramer, Hans van der Vorst, Wiebe Heitman, Jan van Hellemond, Alexander Verburgt and Roddy Chishick for sharing their experience and knowledge.

Many thanks to students from my group for unforgettable two years at TUDelft, especially to: Maryia, Leszek, Samira, Sebastian, Miranda, Joe, Asia, Catalin, Tina, Scrat, Kasia S, Kasia K. and Kasia J.

And finally, never enough thanks to Magda for her constant support.

Łukasz Bednara
Delft, the Netherlands
July 25, 2008

Contents

List of Figures

Chapter 1

Introduction

When designing a new production plant a key aspect is to analyze and predict the production availability. Production systems may be interpreted as the systems that convert raw product (for instance crude oil) into finished product (i.e. gasoline) in a continuous process. The production systems usually consist of a number of connected components such as pumps, compressors, tanks which are required to perform certain functions. Components fail occasionally sometimes in a fully random process, sometimes in a process governed by ageing. The impact of the components on the production depends on the structure of the system and criticality of the failure. The relation between component reliability and system reliability is important for decisions in the design as well as in the operation phase. For instance in the design phase, before the system is being built, one can compare different configurations of the components and check the impact on the system as a whole in order to choose the optimal design structure. In the operation phase information about the unreliable components may help in optimizing maintenance strategies.

In practice the analysis starts from transforming the process flow to the reliability block diagrams (RBD) that represent the logical structure of the system. Figure 1.1 illustrates an example process flow scheme and corresponding reliability block diagram for sub-sea water injection system. The next step is to gather specific characteristics about each component. The most relevant information is the expected functioning time without a failure and also the information about the repair time. Since in most of the situations those quantities are uncertain they have to be modeled by random variables. The choice of proper distributions is very often difficult, especially when data is lacking or scarce. Once the distributions and parameters are known one can predict the performance of each component separately and then combine this information, taking into account the structure of the system, in order to assess the performance of the system.

One of the most important performance measures of repairable units is the availability since it takes into account both failure and repair information. Availability is defined as the probability that an item is available for operation at a specified time

$$A(t) = P(\text{component is available at time } t)$$

An alternative interpretation of $A(t)$ is that if we have a very large number n of identical and independent components that have been put into operation at the same time, $n \cdot A(t)$ is the expected number of components that will be functioning at time t. The availability of the system is defined in the same manner. The availability of the component (or system) depends on its initial condition and both the failure and repair distributions specified for modeling its failure and repair processes. It is a common practice to assume that the system is initially in a good condition and that it is as good as new. In this thesis, however, we consider a more general situation where not necessarily new components at the initial state are also allowed.

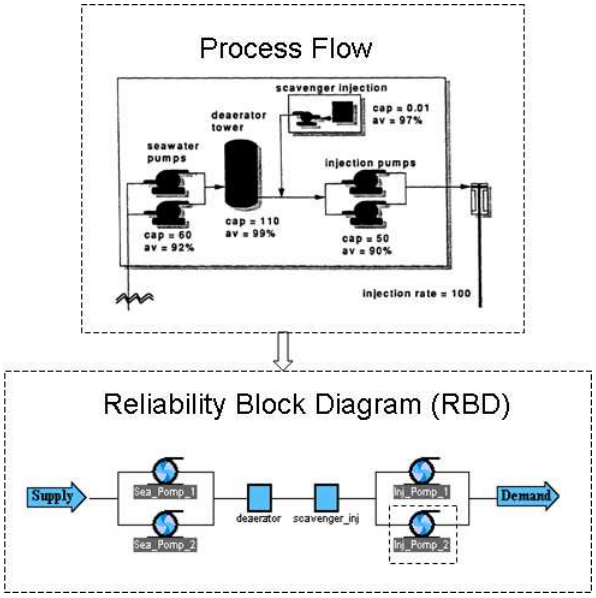


Figure 1.1: From process flow scheme to reliability block diagram: sub-sea water injection system.

1.1 Objective of the Thesis

The aim of this project was to come up with an accurate, robust and fast method for computing different renewal and availability functions on a component level. Particular emphasis was placed on the realistic choices of parameters for life time and repair distributions.

It has to be stressed that there was no intention to analyze the availability on a system level and also there was no statistical analysis regarding the proper choice of the distributions and their parameters.

1.2 Outline of Thesis

This document is structured as follows. Chapter 2 presents basic definitions from renewal theory and also introduces notation for specific renewal and availability functions. Here also the equations for the functions are derived and the analytical solution

for exponential case is given. Chapter 3 is the main part of the thesis and it presents the process of selection of the best method for computing the functions introduced in Chapter 2. The next chapter deals with problem of ”grace periods” and can also be read separately. The conclusions and recommendations are presented in Chapter 5

Chapter 2

Description of the problem

In this chapter we introduce the mathematical statement of the problem and the notation that will be used in next chapters. This chapter is organized as follows. Section 2.1 gives the basic definitions related to the general renewal theory. In Section 2.2 we derive possible equations for the renewal functions of two renewal processes: ordinary and delayed. Next, the four different availability functions are introduced and the corresponding equations are derived.

2.1 Introduction and Notation

Renewal theory is a key tool in reliability modeling. We will now summarize the most important ideas and results of the renewal theory that are needed for later reference. Most of the general theory is given in [41, 15].

Definition 2.1.1. Given the sequence $(T_j : j \geq 1,2,\dots)$ of positive random variables let S_n be the sequence of partial sums

$$S_n = T_1 + \cdots + T_n \ , \ n \geq 1$$

with the convention $S_0 = 0$. The process $\{N(t), t \geq 0\}$ is called a *counting process* associated to $(T_j : j \geq 1,2,\dots)$ if:

$$N(t) = k \iff S_k \leq t < S_{k+1} \tag{2.1}$$

In relation with definition 2.1.1 there are two processes that are of particular interest.

Definition 2.1.2. Let the variables in the sequence $(T_j : j \geq 1,2,\dots)$ be independent and identically distributed then $N(t)$ defined in 2.1.1 is called an *ordinary renewal process* or shortly *renewal process*.

Definition 2.1.3. Let the variables in the sequence $(T_j : j \geq 1,2,\dots)$ be independent, but suppose only $T_2, T_3, \dots$ are identically distributed with distribution function F_T, while T_1 has possibly a different distribution function G, then $N(t)$ defined in 2.1.1 is called a *delayed renewal process* and denoted by $N^D(t)$.

The latter process is a generalization of an ordinary renewal process, that is for $G = F$ the process $N^D(t)$ becomes $N(t)$. The principal objective of renewal theory is to derive properties of certain variables associated with $N(t)$. In this thesis it is of significance and relevance to compute the expected number of renewals in the time interval $(0,t]$.

Definition 2.1.4. *The renewal function* is defined as

$$M(t) = E[N(t)] \text{ and } M^D(t) = E[N^D(t)]$$

Definition 2.1.5. For differentiable renewal function M we define the *renewal density* as

$$m(t) = \frac{dM(t)}{dt}$$

In reliability applications renewal density is sometimes called *the rate of occurrence of failures* or shortly ROCOF and it may be regarded as the mean number of failures per time unit at time t or as the frequency of a unit failures at time t. Thus if $m(t)$ is increasing then the component is deteriorating. On the other hand if $m(t)$ is decreasing the component is improving.
From Definition 2.1.5 it follows that

$$M(t) = \int_0^t m(t)dt \tag{2.2}$$

Definition 2.1.6. Let X and Y be independent life lengths with distribution functions F_X and F_Y and corresponding densities f_X and f_Y. Then the distribution of the sum $X+Y$ is called the *convolution* of F_X and F_Y, and is defined by

$$F_{X+Y}(t) = \int_0^t F_X(t-x)dF_Y(x) = \int_0^t F_Y(t-x)dF_X(x) \tag{2.3}$$

And similarly in case of densities

$$f_{X+Y}(t) = \int_0^t f_X(t-x)\,f_Y(x)dx = \int_0^t f_Y(t-x)\,f_X(x)dx \tag{2.4}$$

The convolution of two functions is denoted by '$\circledast$': $F_{X+Y}(t) = F_X \circledast F_Y(t)$ and $f_{X+Y}(t) = f_X \circledast f_Y(t)$.

Remark 2.1.7. The convolution defined in (2.3) is also called Riemann-Stieltjes convolution and is not commutative for functions that are not equal zero at $t = 0$. The convolution defined by (2.4) is always commutative (it can be shown by simple change of variables). Let F and G be two functions defined on $[0,\infty)$ then

$$\begin{aligned}\int_0^t F(t-x)dG(x) &= \int_0^t F(t-x)G'(x)dx = \int_0^t F(x)G'(t-x)dx \\ &= \int_0^t F'(x)G(t-x)dx - (F(t)G(0) - F(0)G(t)) \\ &= \int_0^t G(t-x)dF(x) - (F(t)G(0) - F(0)G(t))\end{aligned}$$

where in the second line the integration by parts was used. Generalization of the convolution to n identically distributed random variables is straightforward.

Definition 2.1.8. If $T_1, \ldots, T_n$ are iid with common distribution F_T then distribution of the sum $T_1 + \cdots + T_n$ is called *n-fold convolution* of F_T and is defined by

$$F^{(n)}(t) = \int_0^t F^{(n-1)}(t-x)dF_T(x) \tag{2.5}$$

with $F^{(1)}(t) = F_T(t)$. And similarly in case of densities

$$f^{(n)}(t) = \int_0^t f^{(n-1)}(t-x)\, f_T(x)dx \tag{2.6}$$

with $f^{(1)}(t) = f_T(t)$.

In case when T_1 has a different distribution function, say G, then

$$F^{(n)}(t) = G \circledast F^{(n-1)}(t)$$

2.2 Renewal functions

From definition 2.1.1 it follows that

$$N(t) = \sum_{n=1}^{\infty} \mathbf{1}_{\{S_n \le t\}}$$

where $\mathbf{1}_{\{A\}}$ denotes indicator function of the set A. Therefore the renewal function for ordinary renewal process becomes

$$M(t) = E[N(t)] = \sum_{n=1}^{\infty} P(S_n \le t) = \sum_{n=1}^{\infty} F_T^{(n)}(t) \tag{2.7}$$

We next show that the renewal function $M(t)$ satisfies the equation

$$M(t) = F_T(t) + \int_0^t M(t-x)dF_T(x) \tag{2.8}$$

or in convolution notation

$$M(t) = F_T(t) + F_T \circledast M(t)$$

Using directly formula (2.7) we get

$$\begin{aligned} M(t) &= \sum_{n=1}^{\infty} F_T^{(n)}(t) = F_T(t) + \sum_{n=2}^{\infty} F_T^{(n)}(t) \\ &= F_T(t) + \sum_{n=1}^{\infty} F_T^{(n+1)}(t) = F_T(t) + \sum_{n=1}^{\infty} F_T \circledast F_T^{(n)}(t) \\ &= F_T(t) + F_T \circledast \sum_{n=1}^{\infty} F_T^{(n)}(t) \\ &= F_T(t) + F_T \circledast M(t) \end{aligned}$$

and relation (2.8) is established. This equation is known as the *fundamental renewal equation* and sometimes may be solved for $M(t)$.

Immediately by differentiation of (2.8) we obtain similar integral equation for the renewal density $m(t)$

$$m(t) = \frac{dM(t)}{dt} = f_T(t) + \frac{d}{dt}\left(\int_0^t M(t-x)\, f_T(x)dx\right) \tag{2.9}$$

$$= f_T(t) + \int_0^t \frac{d}{dt}\left(M(t-x)\, f_T(x)\right) dx + M(0)\, f_T(t) \tag{2.10}$$

$$= f_T(t) + \int_0^t m(t-x)\, f_T(x)dx + M(0)\, f_T(t) \tag{2.11}$$

(see also Remark 2.2.1) and including that $M(0) = 0$, by its definition, it follows

$$m(t) = f_T(t) + \int_0^t m(t-x)\, f_T(x)dx \tag{2.12}$$

or

$$m(t) = f_T(t) + \int_0^t m(t-x)dF_T(x) \tag{2.13}$$

Remark 2.2.1. Alternative argument may be given in terms of Laplace transforms. Define

$$h(t) = \int_0^t M(t-x)\, f_T(x)dx$$

Then $\mathcal{L}[h(t)] = \mathcal{L}[M(t)] \cdot \mathcal{L}[f_T(t)]$. Since $h(0) = 0$, the Laplace transform of the derivative h' is

$$\begin{aligned}\mathcal{L}\left[h'(t)\right] &= s\mathcal{L}[M(t)] \cdot \mathcal{L}[f_T(t)] = \left(s\mathcal{L}[M(t)] - M(0)\right)\mathcal{L}[f_T(t)] + M(0)\mathcal{L}[f_T(t)] \\ &= \mathcal{L}\left[M'(t)\right]\mathcal{L}[f_T(t)] + M(0)\mathcal{L}[f_T(t)]\end{aligned}$$

So

$$h'(t) = \int_0^t m(t-x)\, f_T(x)dx + M(0)\, f_T(t)$$

and relation (2.11) is established.

Note also that by differentiating (2.7) the renewal density can also be expressed as

$$m(t) = \sum_{n=1}^{\infty} f_T^{(n)}(t) \tag{2.14}$$

where $f_T^{(n)}$ is n-fold convolution of the renewal density. Renewal function may be also expressed in terms of Laplace transform of underlying distribution. According to Appendix B the Laplace transform of (2.8) is

$$M^*(s) = F_T^*(s) + sM^*(s) \cdot F_T^*(s)$$

Solving for $M^*(s)$ gives

$$M^*(s) = \frac{F_T^*(s)}{1 - sF_T^*(s)} \tag{2.15}$$

or

$$M^*(s) = \frac{f_T^*(s)}{s(1 - f_T^*(s))} \tag{2.16}$$

where in last expression we used $F_T^*(s) = f_T^*(s)/s$. If the Laplace transform of densities or distributions are known, one can also find the renewal function from

$$M(t) = \mathcal{L}^{-1}\left[M^*(s)\right] = \mathcal{L}^{-1}\left[\frac{F_T^*(s)}{1 - sF_T^*(s)}\right] \tag{2.17}$$

or

$$M(t) = \mathcal{L}^{-1}\left[M^*(s)\right] = \mathcal{L}^{-1}\left[\frac{f_T^*(s)}{s(1 - f_T^*(s))}\right] \tag{2.18}$$

where $\mathcal{L}^{-1}$ is the inverse Laplace operator.
The following theorem will be useful when deriving the integral equations for the renewal function in delayed case and also for the availability functions:

Theorem 2.2.2. *Suppose H is a bounded function. There exists one and only one function r bounded on finite intervals that satisfies*

$$r(t) = H(t) + \int_0^t r(t-x)dF(x)$$

This function is

$$r(t) = H(t) + \int_0^t H(t-x)dM(x)$$

where $M(t) = \sum_{n=1}^{\infty} F^{(n)}(t)$ is the renewal function (of the ordinary renewal process).

Proof may be found in e.g. [26]. Repeating the argument from ordinary case it is easy to show that the renewal function $M^D(t)$ satisfies

$$\begin{aligned} M^D(t) &= E\left[\sum_{n=1}^{\infty} \mathbf{1}_{\{S_n \le t\}}\right] = P(T_1 \le t) + \sum_{n=2}^{\infty} P(S_n \le t) \\ &= G(t) + \sum_{n=2}^{\infty} G \circledast F_T^{(n-1)}(t) = G(t) + G \circledast \sum_{n=1}^{\infty} F_T^{(n)}(t) \\ &= G(t) + G \circledast M(t) \\ &= G(t) + \int_0^t G(t-x)dM(x) \end{aligned} \tag{2.19}$$

Based on Theorem 2.2.2 equation (2.19) is the solution of the following integral equation

$$M^D(t) = G(t) + \int_0^t M^D(t-x)dF_T(x) \tag{2.20}$$

Similarly as for the function M one can use Laplace transforms to obtain additional representation for function M. By taking Laplace transforms of both sides of (2.20) and solving for $M^*(s)$ gives

$$M^{D^*}(s) = \frac{G^*(s)}{1 - sF_T^*(s)} \tag{2.21}$$

or

$$M^{D^*}(s) = \frac{g^*(s)}{s(1 - f_T^*(s))} \tag{2.22}$$

where g is the density function of the first renewal time.

2.2.1 Asymptotic expansion for the Renewal Function

The first one who introduced asymptotic representations of the renewal function was W.L. Smith [40]. He showed that linear asymptotic approximation for the renewal function is:

$$M(t) \sim \frac{t}{E[T]} - 1 + \frac{E[T^2]}{2E[T]^2}, \ t \to \infty \tag{2.23}$$

It is also possible to derive more accurate approximations. Assuming that at least the first three moments of the interval distribution are finite ref [10] showed that:

$$\begin{aligned} M(t) &= \frac{t}{E[T]} - 1 + \frac{E[T^2]}{2E[T]^2}\frac{S(t)}{E[T]^2} + \frac{E[T^2]}{E[T]^3}\cdot R(t) \\ &\quad + o(t^{-q}\log(t)), \ t \to \infty \\ R(t) &= \int_t^\infty (1 - F(y))dy \\ S(t) &= -\int_t^\infty R(t)dy \end{aligned} \tag{2.24}$$

where q is the highest-order finite moment of the distribution. This higher order approximation is not very useful in practice since for each value of t one has to compute a double integral of the renewal distribution (in case of availability the distribution F is a convolution and has to be approximated as well). Therefore despite the fact that renewal function converges fairly rapidly to (2.24) due to simplicity of (2.23) we keep it as most practical approximation for large values of t.
The asymptotic value for the renewal function may also be derived in delayed case. Ref [33] showed that

$$M_D(t) \sim \frac{t}{E[T]} + \frac{E[T^2] - 2E[T_1]E[T]}{2E[T]^2}$$

where $T \sim F$ and $T_1 \sim G$.

2.3 Availability Functions

We consider a component that can be put into operation and perform its functions for a random period of time. Due to the process of ageing the component may fail and stop to operate, also for a random period of time. It is assumed that after failure the component is repaired and restored to the state "as good as new" (we do not include imperfect repairs). Therefore the component may be in two states either functioning (1) or being repaired (0). Let the sequence $(U_i, i = 1, 2, \dots)$ denote the successive *up-times* of the component. Assume that that up-times are independent and identically distributed with common distribution function $F_U(t) = P(U_i \leq t)$ and mean $E[U_1] = MTTF$ (mean time to failure). Likewise we assume that the corresponding down-times $(D_i, i = 1, 2, \dots)$ are independent and identically distributed with distribution function $F_D(t) = P(D_i \leq t)$ and mean $E[D_1] = MTTR$ (mean time to repair). We will assume that the initial state of the component can be either 0 or 1.
Although a renewal process is defined by the distributions of its inter-arrival times, it

is more convenient to use a binary indicator variable $X_i(t)$ which takes the value 1 if the component is operating at time t and 0 otherwise, where $i = 0$ if the component is initially down and $i = 1$ when it is initially up. If we define the completed repairs, in the process $X_1(t)$, to be the renewals we obtain an ordinary renewal process with renewal periods $U_i + D_i$ for $i = 1, 2, \ldots$. This process is called *an alternating renewal process* and denoted by $N_{1R}(t)$ (initial state is one and it counts completed repairs). An example of a realization of the process $X_1(t)$ is presented in Figure 2.1.

In order to define another renewal process related to $X_i(t)$ we need to assume that

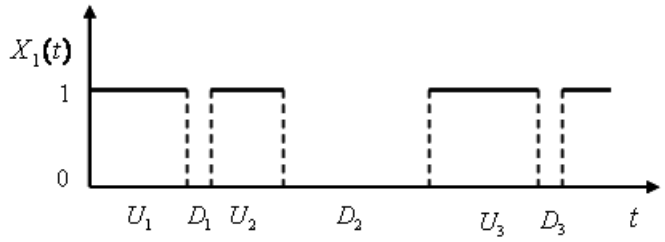


Figure 2.1: Example realization of the process $X_1(t)$

the variables U_i and D_i are also independent for all i. This assumption, usually, is not realistic (see also Section 4.3) however it is crucial in deriving equations for the availability functions. Now if we define the moments the failures occur, in the process $X_1(t)$, to be the renewal moments we obtain a delayed renewal process with renewal periods $U_i + D_{i-1}$ for $i = 2, 3, \ldots$ and first renewal equal to U_1. This renewal process is denoted by N_{1F} (initial state is one and it counts failures). In a similar way we define the processes in case when the initial state is 0, see also Figure 2.3. Let $N_{0R}(t)$ denote the counting process associated to completed repairs in the process $X_0(t)$, and also let $N_{0F}(t)$ denote the process that counts failures in the process $X_0(t)$.

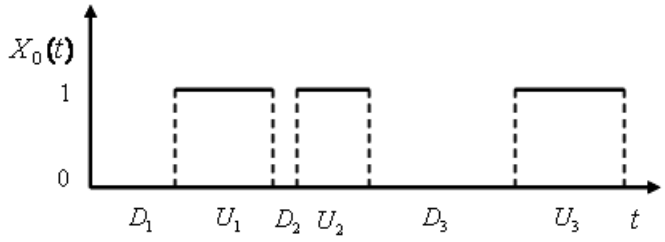


Figure 2.2: Example realization of the process $X_0(t)$

Definition 2.3.1. The *point availability* or *time dependent availability* at time t of a component is defined as the probability that the component is functioning at time t

$$A_i(t) = P(X_i(t) = 1) = E[X_i(t)], \;\; i = 0, 1 \tag{2.25}$$

The existence of this limit follows from an application of the Basic Renewal Theorem 2.2.2 to A_i, assuming that all functions $H(t)$ (see Table 2.4) are directly Riemann integrable. Other availability measures:

- The *limiting availability* or *steady-state availability A* of the component is defined by
$$A = \lim_{t\to\infty} A_i(t), \quad i = 0,1 \tag{2.26}$$

- The *average availability* or *interval availability* of the component in the time interval (t_1,t_2) is defined by
$$A^i_{av}(t_1,t_2) = \frac{1}{t_2-t_1}\int_{t_1}^{t_2} A_i(t)dt \tag{2.27}$$
It can be interpreted as the expected fraction of a given interval of time that the component will be able to operate.

- The *interval reliability* or *mission reliability* is defined as the probability that at specified time time t, the component is operating and will continue to operate for an interval of duration x:
$$R_i(t,x) = P(X_i(u) = 1, t \le u \le t+x)$$
(Note that in case of an ordinary renewal process this quantity corresponds to $P(\gamma(t) > x)$), where $\gamma(t)$ is the remaining lifetime of the item operating at time t. For a clear explanation of the variable $\gamma(t)$ see [26].)

- The *joint availability*, see [6], is the probability that the component is functioning at times t_1 and t_2
$$A_i(t_1,t_2) = P(X_i(t_1) = 1, X_i(t_2) = 1)$$

- The *joint interval availability*, see [6], is the probability that the component is functioning in the intervals $[t_1,t_1+x_1]$ and $[t_2,t_2+x_2]$
$$R_i(t_1,x_1,t_2,x_2) = P(X_i(u) = 1, t_1 \le u \le t_1+x_1, X_i(z) = 1, t_2 \le z \le t_2+x_2)$$

The latter two quantities can be expressed in terms of two availability functions $A_1(t)$, $A_0(t)$ and $R(t,x)$. The interval reliability $R(t,x)$ may be found by solving an integral equation of renewal type. Due to the fact that the point availability functions occur directly or indirectly in all quantities mentioned above (except the interval reliability) they are the most relevant in availability studies.
However complete analysis of the availability requires also two additional functions.

2.3.1 The reason for four different availability functions

From the practical point of view it is important to predict the component's future performance based on its present condition. Let us consider an example.

Example 2.3.2. *Suppose that in a plant there are four identical and independent components that had been put into operation two months ago. Within this time there were repeated cycles of failures and repairs of each component. At this moment in time the components are in the following conditions:*

Case 1: Repair of the first component has just finished

Case 2: Second component has just failed

Case 3: Third component is up and running for last x hours

Case 4: Fourth component is in repair for last y days

The situation is presented in Figure 2.3. Obviously, the fact that the components

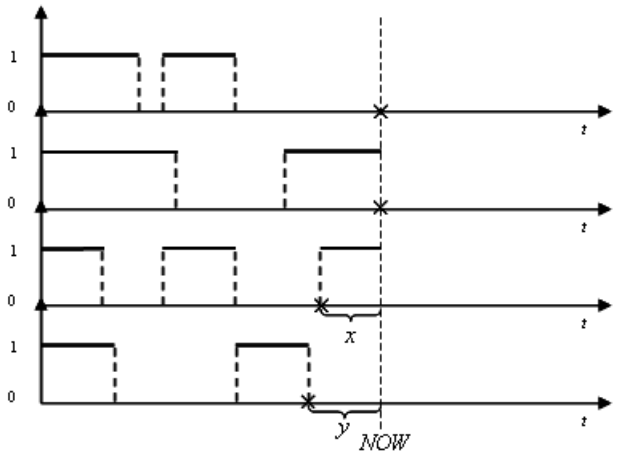


Figure 2.3: States of the components from Example 2.3.2. Components are labeled starting from the top. Crosses denote the time of the last event.

started functioning at the same tame does not have any influence on the prediction because each repair/replacement brings the component to the state as good as new. The only important information is that when the last event (failure or completed repair) occurred.
We are interested in the availability functions for all 4 components.

Note that the cases presented in the example are complete and describe any possible situation that may occur. The reader immediately realizes that Cases 1 and 2 correspond to the processes $X_1(t)$ and $X_0(t)$ respectively defined above. Therefore the future performance of the components one and two may be measured by the availability functions $A_1(t)$ and $A_0(t)$. Cases 3 and 4 are delayed versions of cases 1 and 2 and will be considered in details in separate sections where we derive equations for delayed equivalents of $A_1(t)$ and $A_0(t)$.
Nevertheless it is clear that four availability functions associated with the cases from Example 2.3.2 will cover all possible situations that may occur in practice. In most of the situations the modeling is done starting at any time t and very often it is not the time of the event therefore the delayed availability functions are very important.

2.3.2 The function A_1

Recall that the availability function A_1 is defined as

$$A_1(t) = P(X_1(t) = 1).$$

The processes $N_{1R}(t)$ and $N_{1F}(t)$ count the number of completed repairs and number of failures respectively. Let $M_{1R}(t)$ and $M_{1F}(t)$ be the renewal functions of those processes, i.e.

$$M_{1R}(t) = E[N_{1R}(t)] \text{ and } M_{1F}(t) = E[N_{1F}(t)] \tag{2.28}$$

Note that the number of repairs equals the number of failures when the component is functioning, see Figure 2.1. On the other hand when the component is down the number of failures equals the number of repairs plus one. Thus

$$\begin{aligned} X_1(t) = 1 &\iff N_{1R}(t) = N_{1F}(t) \\ X_1(t) = 0 &\iff N_{1R}(t) + 1 = N_{1F}(t) \end{aligned}$$

Using above equivalences we can easily derive the relation between the renewal functions and the availability function. Namely

$$\begin{aligned} M_{1F}(t) - M_{1R}(t) &= E[N_{1F}(t) - N_{1R}(t)] \\ &= E[N_{1F}(t) - N_{1R}(t) | X_1(t) = 1] P(X_1(t) = 1) \\ &\quad + E[N_{1F}(t) - N_{1R}(t) | X_1(t) = 0] P(X_1(t) = 0) \\ &= 1 - P(X_1(t) = 1) \end{aligned}$$

Therefore

$$A_1(t) = M_{1R}(t) - M_{1F}(t) + 1 \tag{2.29}$$

Now we present how to obtain an integral equation for A_1. Function $M_{1R}(t)$ corresponds to the ordinary renewal process N_{1R} that has renewals defined to be the sum $U_i + D_i$ for $i = 1, 2, \ldots$. Thus the underlying distribution of this process is the convolution of F_U and F_D (see Definition 2.1.6)

$$F_{U+D}(t) = F_U \circledast F_D(t)$$

From (2.8) it follows that $M_{1R}(t)$ satisfies the following integral equation

$$M_{1R}(t) = F_{U+D}(t) + \int_0^t M_{1R}(t-x) dF_{U+D}(x) \tag{2.30}$$

On the other hand function $M_{1F}(t)$ corresponds to the delayed renewal process N_{1F} that has renewals defined to be the sum $U_i + D_{i-1}$ for $i = 2, 3, \ldots$ with first renewal equal to U_1. Since the variables U_i and D_{i-1} for $i = 2, 3, \ldots$ are independent with distributions F_U and F_D respectively the underlying distribution of the process is the convolution of F_U and F_D and is the same as in previous case. The distribution of the first arrival is just the distribution the of up-time F_U.
From (2.20) it follows that $M_{1F}(t)$ satisfies the following integral equation

$$M_{1F}(t) = F_U(t) + \int_0^t M_{1F}(t-x) dF_{U+D}(x) \tag{2.31}$$

Subtracting (2.31) from (2.30) gives

$$M_{1R}(t) - M_{1F}(t) = F_{U+D}(t) - F_U(t) + \int_0^t \left(M_{1R}(t-x) - M_{1R}(t-x) \right) dF_{U+D}(x)$$

or equivalently

$$A_1(t) = 1 + F_{U+D}(t) - F_U(t) + \int_0^t (A_1(t-x)+1)\, dF_{U+D}(x).$$

Therefore

$$A_1(t) = 1 - F_U(t) + \int_0^t A_1(t-x) dF_{U+D}(x) \tag{2.32}$$

As a result of Theorem 2.2.2 the solution of (2.32) is

$$A_1(t) = 1 - F_U(t) + \int_0^t (1 - F_U(t-x))\, dM_{1R}(x). \tag{2.33}$$

The function A_1 can also be expressed in terms of Laplace transform. It is straightforward to show by taking Laplace transforms on both sides of (2.32) and solving for $A_1^*(s)$ that

$$A_1^*(s) = \frac{1 - f_U^*(s)}{s(1 - f_{U+D}^*(s))} \tag{2.34}$$

2.3.3 The function A_0

The derivation of equations for the function A_0 is almost exactly the same as in case of A_1. Recall that availability function A_0 is defined as

$$A_0(t) = P(X_0(t) = 1)$$

The processes $N_{0R}(t)$ and $N_{0F}(t)$ count the number of completed repairs and number of failures respectively. Let $M_{0R}(t)$ and $M_{0F}(t)$ be the renewal functions of those processes, i.e.

$$M_{0R}(t) = E[N_{0R}(t)] \text{ and } M_{0F}(t) = E[N_{0F}(t)] \tag{2.35}$$

Note that the number of repairs equals the number of failures when the component is down, see Figure 2.3. On the other hand when the component is functioning the number of repairs equals the number of failures plus one. Thus

$$\begin{aligned} X_0(t) = 0 &\iff N_{0R}(t) = N_{0F}(t) \\ X_0(t) = 1 &\iff N_{0R}(t) = N_{0F}(t) + 1 \end{aligned}$$

Using above equivalences it is easy to derive the relation between the renewal functions and the availability function. Namely

$$\begin{aligned} M_{0R}(t) - M_{0F}(t) &= E[N_{0R}(t) - N_{0F}(t)] \\ &= E[N_{0R}(t) - N_{0F}(t) | X_0(t) = 1] P(X_0(t) = 1) \\ &\quad + E[N_{0R}(t) - N_{0F}(t) | X_0(t) = 0] P(X_0(t) = 0) \\ &= P(X_0(t) = 1) \end{aligned}$$

Therefore

$$A_0(t) = M_{0R}(t) - M_{0F}(t) \tag{2.36}$$

Similarly as in previous case we present how to obtain an integral equation for A_0. Function $M_{0F}(t)$ corresponds to the ordinary renewal process (N_{0R}) that has renewals defined to be the sum $D_i + U_i$ for $i = 1, 2, \ldots$. Thus the underlying distribution of this process is the convolution of F_D and F_U (see Definition 2.1.6)

$$F_{U+D}(t) = F_U \circledast F_D(t)$$

Since the renewal process is completely determined by the distribution of the inter-arrivals it follows that

$$N_{0F}(t) \stackrel{d}{=} N_{1R}(t)$$

Therefore

$$M_{0F}(t) = M_{1R}(t)$$

However for completeness we also write an integral equation for $M_{0F}(t)$

$$M_{0F}(t) = F_{U+D}(t) + \int_0^t M_{0F}(t-x) dF_{U+D}(x) \tag{2.37}$$

On the other hand function $M_{0R}(t)$ corresponds to the delayed renewal process N_{0R} that has renewals defined to be the sum $U_{i-1} + D_i$ for $i = 2, 3, \ldots$ with first renewal equal to D_1. Since the variables U_{i-1} and D_i for $i = 2, 3, \ldots$ are independent with distributions F_U and F_D respectively the underlying distribution of the process is the convolution of F_U and F_D and is the same as in previous case. The distribution of the first arrival is just the distribution of the down-time F_D.
From (2.20) it follows that $M_{0R}(t)$ satisfies the following integral equation

$$M_{0R}(t) = F_D(t) + \int_0^t M_{0R}(t-x) dF_{U+D}(x) \tag{2.38}$$

Subtracting (2.37) from (2.38) gives

$$M_{0R}(t) - M_{0F}(t) = F_D(t) - F_{U+D}(t) + \int_0^t \left(M_{0R}(t-x) - M_{0F}(t-x)\right) dF_{U+D}(x)$$

Therefore

$$A_0(t) = F_D(t) - F_{U+D}(t) + \int_0^t A_0(t-x) dF_{U+D}(x) \tag{2.39}$$

As a result of Theorem 2.2.2 the solution of (2.39) is

$$A_0(t) = F_D(t) - F_{U+D}(t) + \int_0^t \left(F_D(t-x) - F_{U+D}(t-x)\right) dM_{0F}(x) \tag{2.40}$$

The function A_0 can also be expressed in terms of Laplace transform. It is straightforward to show that by taking Laplace transforms on both sides of (2.39) and solving for $A_0^*(s)$ gives

$$A_0^*(s) = \frac{f_D^*(s) - f_{U+D}^*(s)}{s(1 - f_{U+D}^*(s))} \tag{2.41}$$

To model the availability when the component is not new (see Case 3 in Example 2.3.2) or when the repair is in progress (see Case 4 in Example 2.3.2) we need to define new state variables. If the component has been functioning for x time units then its remaining lifetime has a different distribution than the lifetime of a new component of the same type. For the exponential case it will make no difference since this distribution possesses so called *memoryless property*. However in general, the distribution of the remaining lifetime is a conditional distribution. Similarly with the repair time. It is not the purpose now to specify this distribution, but it is important to emphasize that the distribution of the time to the first event may differ from the distributions of next up and down events. Therefore it is convenient to call these cases delayed versions of cases 1 and 2 from Example 2.3.2. We introduce the following notation.
The cases when the component is initially up and initially down are considered separately.
The system is initially up. Let $R, U_2, U_3, \ldots$ be the sequence of independent up-times. Similarly as before assume that variables $U_2, U_3, \ldots$ are identically distributed with common distribution function $F_U(t) = P(U_i \leq t)$. Let $F_R(t)$ denote the cumulative distribution function of the first up time R. The sequence of down times $D_1, D_2, D_3, \ldots$ remains unchanged. Let $X_1^D(t)$ denote the state of the component at time t (an example of a realization of the process $X_1^D(t)$ is presented in Figure 2.4).

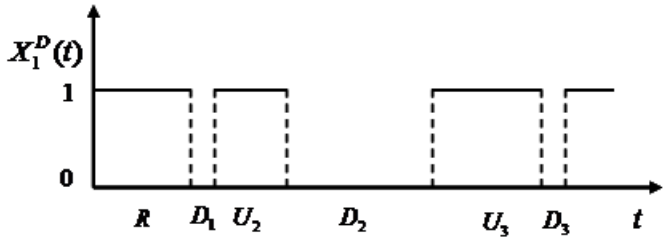


Figure 2.4: Example realization of the process $X_1^D(t)$

The system is initially down. Let $S, D_2, D_3, \ldots$ be the sequence of independent down-times. Similarly as before assume that variables $D_2, D_3, \ldots$ are identically distributed with common distribution function $F_D(t) = P(D_i \leq t)$. Let $F_S(t)$ denote the cumulative distribution function of the first down time S. The sequence of up times $U_1, U_2, U_3, \ldots$ remains unchanged. Let $X_0^D(t)$ denote the state of the component at time t (an example of a realization of the process $X_0^D(t)$ is presented in Figure 2.5).

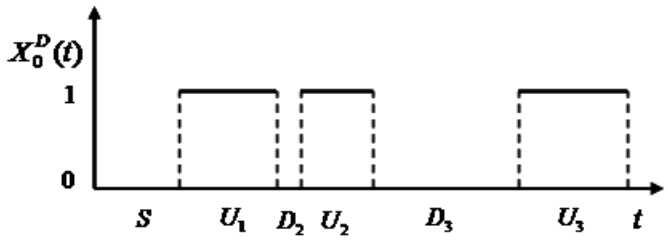


Figure 2.5: Example realization of the process $X_0^D(t)$

2.3.4 The function A_1^D

Let A_1^D be the availability function related to the process $X_1^D(t)$, namely

$$A_1^D(t) = P(X_1^D(t) = 1)$$

In a similar manner as in previous sections we define an associated processes. Let $N_{1R}^D(t)$ be the renewal process that counts completed repairs in process $X_1^D(t)$ and similarly let $N_{1F}^D(t)$ be the renewal process that counts failures. The renewal functions of these processes are denoted as

$$M_{1R}^D(t) = E[N_{1R}^D(t)] \text{ and } M_{1F}^D(t) = E[N_{1F}^D(t)] \tag{2.42}$$

Since the renewal processes $N_{1R}^D(t)$ and $N_{1F}^D(t)$ are equivalent respectively to $N_{1R}(t)$ and $N_{1F}(t)$ (they count the same events) one can use the same reasoning as in Section 2.4 to derive the relation between the availability function and the renewal functions. Therefore

$$A_1^D(t) = M_{1R}^D(t) - M_{1F}^D(t) + 1 \tag{2.43}$$

Since the distribution of the first lifetime is different from any other distribution in the model both processes $N_{1R}^D(t)$ and $N_{1F}^D(t)$ are delayed (see also Figure 2.4). The first distribution of the process $N_{1R}^D(t)$ is the convolution of F_R and F_D. Next renewals are formed by the sequence $U_i + D_i$, $i = 1, 2, \ldots$. Thus the underlying distribution of this process is the convolution of F_U and F_D. In the same manner we obtain that the first renewal in the process $N_{1F}^D(t)$ has the distribution F_R and the underlying distribution is also the convolution of F_U and F_D.

Now we present how to obtain the integral equation for A_1^D. Since both renewal functions correspond to the delayed renewal processes from (2.20) it follows that:

$$M_{1R}^D(t) = F_{R+D}(t) + \int_0^t M_{1R}^D(t-x)dF_{U+D}(x) \tag{2.44}$$

and

$$M_{1F}^D(t) = F_R(t) + \int_0^t M_{1F}^D(t-x)dF_{U+D}(x) \tag{2.45}$$

Subtracting (2.44) from (2.45) we obtain that the availability function $A_1^D(t)$ satisfies the following integral equation

$$A_1^D(t) = 1 - F_R(t) + F_{R+D}(t) - F_{U+D}(t) + \int_0^t A_1^D(t-x)dF_{U+D}(x) \tag{2.46}$$

Note that if the first up-time has the same distribution as the next ones , i.e. $F_R(t) = F_T(t)$, then equation (2.46) reduces to the equation (2.32) for the availability function A_1.

Function A_1^D can also be expressed in terms of Laplace transforms. It is straightforward to show that by taking Laplace transforms on both sides of (2.46) and solving for $A_1^{D^*}(s)$ that

$$A_1^{D^*}(s) = \frac{1 - f_R^*(s) + f_{R+D}^*(s) - f_{U+D}^*(s)}{s(1 - f_{U+D}^*(s))} \tag{2.47}$$

2.3.5 The function A_0^D

Let A_0^D denote the availability function of the process $X_0^D(t)$

$$A_0^D(t) = P(X_0^D(t) = 1).$$

Similarly as in previous sections we define the renewal process that counts the number of completed repairs and number of failures. Let $N_{0R}^D(t)$ be the renewal process that counts completed repairs in process $X_0^D(t)$ and similarly let $N_{0F}^D(t)$ be the renewal process that counts failures. The renewal functions of these processes are denoted as

$$M_{0R}^D(t) = E[N_{0R}^D(t)] \text{ and } M_{0F}^D(t) = E[N_{0F}^D(t)] \tag{2.48}$$

Since the renewal processes $N_{0R}^D(t)$ and $N_{0F}^D(t)$ are equivalent respectively to $N_{0R}(t)$ and $N_{0F}(t)$ (they count the same events) one can use the same reasoning as in Section 2.3.3 to derive the relation between the availability function and the renewal functions. Therefore

$$A_0^D(t) = M_{0R}^D(t) - M_{0F}^D(t) \tag{2.49}$$

Since the distribution of the first downtime is different than any other distribution in the model both processes $N_{0R}^D(t)$ and $N_{0F}^D(t)$ are delayed (see also Figure 2.5). The first distribution of the process $N_{0F}^D(t)$ is the convolution of F_S and F_U. Next renewals are formed by the sequence $D_i + U_i$, $i = 1, 2, \ldots$. Thus the underlying distribution of this process is the convolution of F_U and F_D. In the same manner we obtain that the first renewal in the process $N_{0R}^D(t)$ has the distribution F_S and the underlying distribution is also the convolution of F_U and F_D.
Now we present how to obtain the integral equation for A_0^D. Since both renewal functions correspond to the delayed renewal processes from (2.20) it follows that:

$$M_{0R}^D(t) = F_S(t) + \int_0^t M_{0R}^D(t-x) dF_{U+D}(x) \tag{2.50}$$

and

$$M_{0F}^D(t) = F_{S+U}(t) + \int_0^t M_{0F}^D(t-x) dF_{U+D}(x) \tag{2.51}$$

Subtracting (2.50) from (2.52) we obtain that the availability function $A_0^D(t)$ satisfies the following integral equation

$$A_0^D(t) = F_S(t) - F_{S+U}(t) + \int_0^t A_0^D(t-x) dF_{U+D}(x) \tag{2.52}$$

Note that if the first down-time has the same distribution as the next ones , i.e. $F_S(t) = F_D(t)$, then equation (2.52) reduces to equation (2.39) for the availability function A_0.
Function A_1^D can also be expressed in terms of Laplace transforms. It is straightforward to show that by taking Laplace transforms on both sides of (2.32) and solving for $A_0^{D^*}(s)$ gives

$$A_0^{D^*}(s) = \frac{f_S^*(s) - f_{S+U}^*(s)}{s(1 - f_{U+D}^*(s))} \tag{2.53}$$

2.3.6 Planned Maintenance with variable duration – function A_{PM}

This section presents an additional application of the function A_0^D that is related to planned maintenance. Let us assume that there is a planned maintenance scheduled at fixed time t_{PM} in the future. The choice of t_{PM} is determined based on the maintenance strategy (for instance block replacement, age replacement, condition-based maintenance etc.) and it is not of our interest. For the overview of the maintenance strategies and optimal choice of t_{PM} (maintenance optimization) we refer to [?]. Usually planned maintenance actions have constant duration, however there are some situations in which the time to complete the maintenance is not known upfront and has to be modeled by a random variable, say S. Additionally assume that the maintenance is carried out regardless of the state of a component at time t_{PM} (it can be up or down). After the maintenance is completed a component is subject again to its own rhythm of lifetimes (U_i) and repair times (D_i). It should be emphasized that the distribution of the planned maintenance S may be different from the distribution of usual repair time R_i. This situation corresponds to the process $X_0^D(t)$ introduced above: it starts from down and the time to the first event has a different distribution than the other events. Therefore the renewal and availability functions presented in Section 2.3.5 can be used directly. To distinguish those two cases we denote the availability function for the planned maintenance by $A_{PM}(t)$. (The difference lies only in the interpretation of the first down time. If it represents the remaining repair time, then it corresponds to the function A_0^D, and if it represents planned maintenance then it corresponds to A_{PM}, but from the mathematical point of view both functions satisfy the same equations,as introduced in Section 2.3.5).

2.3.7 Steady state availability

Similarly as in case of the renewal functions the asymptotic behavior of the availability functions is known. The following theorem presents the asymptotic behavior of the the general solution to the renewal integral equation.

Theorem 2.3.3. *(The Basic Renewal Theorem). Let F be the distribution function of a positive random variable with finite mean μ. Suppose that H is directly Riemann integrable and that r is the solution of the renewal equation*

$$r(t) = H(t) + \int_0^t r(t-x)dF(x)$$

Then

$$\lim_{t\to\infty} r(t) = \frac{1}{\mu}\int_0^\infty H(x)dx$$

The *steady state availability* or *limiting* availability is defined as a the asymptotic value for an availability function. It can be show that, steady the state availability is the same for all four availability functions and is equal to

$$A = \lim_{t\to\infty} A_i(t) = \lim_{t\to\infty} A_i^D(t) = \frac{E[U_1]}{E[U_1]+E[D_1]},\ \ i = 1,2 \tag{2.54}$$

This follows from the fact that the initial distribution of the renewal process does not have influence on its asymptotic behavior.

2.4 Analytical solution for the exponential case

Let $F_U(t) = 1 - e^{-\lambda t}$ and $F_D(t) = 1 - e^{-\mu t}$ be the distributions of up-times and down-times respectively. We use technique of Laplace transforms to obtain all renewal and availability functions for the exponential case. For properties of Laplace transforms see Appendix B. For the exponential case it is straightforward to show that the Laplace transforms of the densities of U and D are

$$f_U^*(t) = \frac{\lambda}{\lambda + s} \text{ and } f_D^*(t) = \frac{\mu}{\mu + s}$$

1. **Function** A_1
   Based on equations (2.22) and (2.30) the Laplace transform of M_{1R} is

   $$M_{1R}^*(s) = \frac{f_U^*(t) \cdot f_D^*(t)}{s(1 - f_U^*(t) \cdot f_D^*(t))}$$

   Substituting and simplifying we obtain

   $$M_{1R}^*(s) = \frac{\lambda\mu}{(\lambda + \mu + s)s^2}$$

   Since above representation is a fraction of two polynomials it can be easily inverted. We obtain

   $$M_{1R}(t) = \frac{\lambda\mu}{\lambda + \mu}\left(t - \frac{1 - e^{(\lambda+\mu)t}}{\lambda + \mu}\right) \tag{2.55}$$

   Similarly for the function M_{1F}. Based on the equations (2.16) and (2.31) it is straightforward algebra to show that

   $$M_{1F}^*(s) = \frac{f_U^*(t)}{s(1 - f_U^*(t) \cdot f_D^*(t))} = \frac{(\mu + s)\lambda}{(\lambda + \mu + s)s^2}$$

   Inverting gives

   $$M_{1F}(t) = \frac{\lambda\mu}{\lambda + \mu} \cdot t + \frac{\left(1 - e^{(\lambda+\mu)t}\right)\lambda^2}{(\lambda + \mu)^2} \tag{2.56}$$

   Therefore it follows from the relation (2.29) between the renewal functions and the availability function that

   $$A_1(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} \cdot e^{-(\lambda+\mu)t} \tag{2.57}$$

2. **Function** A_0
   We proceed exactly in the same manner as in the case above. The Laplace transform of M_{0R} is

   $$M_{0R}^*(s) = \frac{f_D^*(t)}{s(1 - f_U^*(t) \cdot f_D^*(t))} = \frac{(\lambda + s)\mu}{(\lambda + \mu + s)s^2}$$

Inverting gives

$$M_{0R}^{*}(s) = \frac{\lambda\mu}{\lambda+\mu}\cdot t + \frac{\left(1-e^{(\lambda+\mu)t}\right)\mu^2}{(\lambda+\mu)^2} \tag{2.58}$$

As it was remarked in Section 2.3.3 $M_{0F}(t) = M_{1R}(t)$ therefore based on relation (2.36) the availability function for the exponential case is

$$A_0(t) = \frac{\mu}{\lambda+\mu} - \frac{\mu}{\lambda+\mu}\cdot e^{-(\lambda+\mu)t} \tag{2.59}$$

3. **Function** A_1^D
   If we assume that the component has the exponential lifetime then the distribution of the remaining lifetime (given that the component is functioning for x time units) is again exponentially distributed with the same parameter λ (due to the memoryless property). Therefore the formulas for the function A_1 apply directly. However, we could consider a case when the distribution of the first life time is exponential but with different parameter, say ζ. But since the case with the distribution that does not represent remaining lifetime does not have any practical application we omit it.

4. **Function** $A_0^D(A_{PM})$
   Similarly as in case of A_1^D if the downtimes are exponentially distributed the remaining repair time has again exponential distribution with the same parameter and formulas for A_0 apply directly. To the contrary with previous situation there is an application of the case when the first down-time has exponential distribution but with a different parameter. As the example let us take planned maintenance which duration is exponentially distributed with $F_S(t) = 1-e^{-\eta t}$, see also Section 2.3.6. We proceed in the same manner as in previous cases (with slightly more complicated inversion of the transforms). Since $f_S^*(t) = \eta/(\eta+s)$ the Laplace transform of M_{0R}^D is

$$M_{0R}^{D\,*}(s) = \frac{f_S^*(t)}{s(1-f_U^*(t)\cdot f_D^*(t))} = \frac{\eta\,(\lambda+s)\,(\mu+s)}{(\eta+s)\,s^2\,(\lambda+\mu+s)}$$

Inverting gives

$$\begin{aligned} M_{0R}^D(t) = \frac{1}{(-\eta+\lambda+\mu)\eta\,(\lambda+\mu)^2}\Big[&3\lambda\mu^2\eta + e^{-\eta t}\,(\mu-\eta)\,(\lambda-\eta)\,(\lambda+\mu)^2 \\ &+ \mu^3\,(-\lambda+\eta+\eta t\lambda) - \mu\left(\mu+\lambda\left(1+t\mu+e^{-(\lambda+\mu)t}\right)\right)\eta^2 \\ &+ (-\mu+\eta+\mu t\eta)\,\lambda^2\,(\lambda+2\mu-\eta)\Big] \end{aligned} \tag{2.60}$$

Similarly for M_{0F}^D the Laplace transform is

$$M_{0F}^{D\,*}(s) = \frac{f_S^*(t)\,f_U^*(t)}{s(1-f_U^*(t)\cdot f_D^*(t))} = \frac{(\mu+s)\,\lambda\eta}{(\lambda+\mu+s)\,s^2\,(\eta+s)}$$

After inverting we obtain

Renewal functions		Availability functions	
$H(t)$	$r(t)$	$H(t)$	$r(t)$
$F_{U+D}(t)$	$M_{1R}(t)$	$1-F_U(t)$	$A_1(t)$
$F_U(t)$	$M_{1F}(t)$		
$F_{U+D}(t)$	$M_{0F}(t)$	$F_D(t)-F_{U+D}(t)$	$A_0(t)$
$F_D(t)$	$M_{0R}(t)$		
$F_{R+D}(t)$	$M_{1R}^D(t)$	$1-F_R(t)+F_{R+D}(t)-F_{U+D}(t)$	$A_1^D(t)$
$F_R(t)$	$M_{1F}^D(t)$		
$F_{S+U}(t)$	$M_{0F}^D(t)$	$F_S(t)-F_{S+U}(t)$	$A_0^D(t)$
$F_S(t)$	$M_{0R}^D(t)$		

Table 2.1: Combination of possible cases for the renewal and availability functions

$$M_{0F}^D(t) = \frac{\lambda^2}{(\lambda+\mu)^2} + \frac{\lambda}{\eta\,(\lambda+\mu)^2\,(-\eta+\lambda+\mu)}\Big[\mu^2\eta + e^{-\eta t}\,(\lambda+\mu)^2\,(\mu-\eta) + \left(-t\mu^2 + \left(e^{-(\lambda+\mu)t} - t\mu\right)\lambda\right)\eta^2 + \mu\left(\lambda\,(2\mu t\eta - 2\mu + \eta) + (\eta t - 1)\left(\mu^2+\lambda^2\right)\right)\Big] \tag{2.61}$$

Therefore, based on relation (2.49), the availability function A_0^D or better A_{PM} for the exponential case is

$$A_{PM}(t) = -\frac{-\lambda e^{-\eta t}\eta + e^{-\eta t}\lambda\mu + \lambda\eta\, e^{-(\lambda+\mu)t} - \lambda\mu - e^{-\eta t}\mu\eta + \mu\eta - \mu^2 + e^{-\eta t}\mu^2}{(-\eta+\lambda+\mu)(\lambda+\mu)} \tag{2.62}$$

Note that all three functions converge to the same limit

$$\lim_{t\to\infty} A_1(t) = \lim_{t\to\infty} A_0(t) = \lim_{t\to\infty} A_{PM}(t) = \frac{\mu}{\mu+\lambda}$$

as it is expected.

The equations for all renewal and availability functions introduced in this chapter can be written in general as the following integral equation

$$r(t) = H(t) + \int_0^t r(t-x)\,dF_{U+D}(x) \tag{2.63}$$

where the function r is unknown and F_{U+D}, H are given (in most of the cases F_{U+D} has to be approximated). The specific choices of H for the renewal and availability functions are gathered in Table 2.4. Equation (2.63) is called *an integral equation of renewal type*.

Chapter 3

Overview of concepts and methods

In previous chapter we have introduced four different availability functions and related renewal functions. The exact analytical solution for those functions may be found only in very simple cases, see for instance exponential case in Section 2.4. Therefore the availability and renewal functions have to approximated by using numerical methods. In this chapter we present the review of the literature that deals with this problem. Although the number of publications that deal with approximation of the renewal functions is rather large there are only few methods that deal with the availability function directly. As it was shown in Chapter 2 all availability functions may be founds via renewal functions therefore the methods for renewal functions are also important. There are two possibilities to find availability function from renewal functions:[1]:

$$A_1(t) = M_{1R}(t) - M_{1F}(t) + 1 \tag{3.1}$$

$$A_1(t) = 1 - F_U(t) + \int_0^t (1 - F_U(t-x))\, dM_{1R}(x) \tag{3.2}$$

The drawback of the first approach is that it requires double effort for computing two renewal functions. The second method is even less efficient while taking into consideration the fact that for each time t, one has to integrate over M_{1R}. For instance let us say that A_1 is required at N grid points, then using (3.1) one needs to compute two renewal functions in total on $2N$ points. Using (3.2) it is even worse and it requires kN points, for some large k. Due to these issues we decide to classify the available methods on those that can compute availability functions and those that can compute renewal functions. The advantage of using the method for availability is that they can also be used for renewal function, opposite relation is not always possible. Therefore based on the available literature review we classify the methods as follows:

- Availability function
  - Direct solution of the renewal equation

[1] This also applies to the other availability functions.

  - Laplace transforms
  - Spline approximation
  - Bounds

- Renewal function
  - All methods for computing availability function
  - Series expansion
  - Approximations

It is not possible to analyze and compare carefully all methods. Before approaching the stage of description and selection of the methods, the requirements that will help in making the decision should be specified.

3.1 The requirements

In this section we present the requirements that are used for selecting the method(s) for computing the availability and renewal functions. Different approaches lead to various difficulties and it is sometimes not possible to select the method that is capable to deal with all possible situations. To exemplify the method of [42] proposed to compute renewal function for Weibull case is quite fast but inappropriate for larger arguments, on the other hand approach suggested by [22] is simple but it requires n^{th} derivative of the distributions. Other problems that may arise are: the error estimation, difficulties with different ranges of parameters or complexity of the algorithm and its implementation. For the propose of this project we decide to choose the following requirements based on which the selection is made:

1. Realistic values of the parameters. In practice we are interested in components that have a good reliability. So either their time to failure is long or their repair time is short (and preferably both). For instance sub-sea equipment is designed to operate without a failure for few decades, but if a serious problem occurs a repair time lasts for few months. On the other hand a pump in a plant may trip a few times per year, but recovery is a matter of hours. In the context of applicability to oil & gas equipment we are interested in components for which

$$MTTF \in [0.1, 100] \text{ years and } MTTR \in [1, 4000] \text{ hours} \tag{3.3}$$

   under the constraint:

$$A = \frac{MTTF}{MTTF + MTTR} \in [0.95, 1] \tag{3.4}$$

   Where A is the limiting availability defined in Chapter 2. Obviously the higher the availability the more valuable the component, hence very often A is close to one. The method should work for as many choices of the parameters as possible.

2. Distributions of interest. In principle any distribution defined for non-negative values can represent life time or repair time. It is hardly probable to find a method that works well for all distributions, and this is not required. But it is

required to find a method that can take as an input a generic distribution and be able to produce some results. The proper choice of the distributions is not easy and it is behind the scope of this report. However a particular emphasis is put on the following realistic distributions:

a) Life time: Weibull distribution with constraint on the shape parameter $\beta \in [0, 32]$.

b) Repair time: exponential distribution, constant, Lognormal distribution

For description and characteristics of those distributions see Appendix A.

3. It is significant to have one method that can compute both renewal and availability functions.

4. Accuracy. This is the key factor and at the same time the most difficult to asses. We require the method to give reasonable results and also it is valuable to have some error estimate. However it is sometimes better to have a method that gives more accurate results and poor error estimation than to have a method that gives poor approximation and good error estimate.

5. Speed. We require the method to be reasonably fast. Practical use of the availability models requires above 100 components (sometimes even 2000). Therefore the speed of calculations for one failure mode is a very important factor. If the computations for one component are slow but acceptable it may appear that the waiting time for the calculations of the whole system of components is not acceptable.

6. The method should work without predefined end point. This requirement is related to the numerical steady-state detection see Section 3.7.3

7. The last but also very important requirement is that the method should be robust on the choices of parameters and relatively simple in implementation.

Next section presents an overview of the available approaches and discusses the most suitable methods for the problem.

3.2 Selection of the methods

3.2.1 Laplace transforms

Laplace transforms are a valuable tool in theoretical study of renewal processes and several algorithms are available for their inversion, (see [1] for an overview). As it was shown in the previous chapter all availability and renewal functions can be expressed in terms of Laplace transforms of up and down time densities. The Laplace transform of Weibull density is expressed in terms of alternating series that converges slowly, for lognormal distribution no analytical expression exists. Therefore a numerical inversion without analytical expressions seem to be difficult. As suggested by [44] one can approximate the density by some polynomial for which the Laplace transform is analytically tractable and then use an algorithm for numerical inversion. Albeit the choice

of the polynomial for the distributions of our interest is slightly difficult. There is also a positive aspect when using Laplace transforms that is valuable for the discussed problem. Recall that repair time usually is much shorter than lifetime. This leads to big differences in densities of up and down time. Choudhury and Whitt [11] proposed a very simple scaling procedure that allows to deal with comparable densities. For a short and clear description of this method we refer to [44]. However, since analytical expressions for the Laplace transforms are not easily accessible, this approach is not continued.

3.2.2 Series expansion

Series expansion approaches were applied only for approximation to a renewal function and none of such methods was found for computing an availability function. Most of the literature is focused on renewal process with Weibull distribution, see [42, 27, 14]. In case of this study renewal distribution is usually a convolution of Weibull distribution with some distribution for repair time therefore those methods do not apply directly. However, in case when down time distribution is constant (thus renewal distribution is a shifted Weibull) the distribution may be scaled so that it is Weibull. This restricts the usage of series expansion methods for only one case. As mentioned by [7] it is easy to note that those series are of little use for numerical work especially with large values of t. This is due to the term with alternating sign. However the series method of Smith and Leadbetter [42] is worth considering in more detail since it converges quickly for $0 < \beta < 1$. If $M(t)$ denote the renewal function with underlying distribution Weibull(α, β) then $M(t)$ can be written as

$$M(t) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1} \cdot A_k \cdot (t/\alpha)^{k\beta}}{\Gamma(k\beta + 1)}$$

where the constants A_k are determined by the following recursive formula

$$A_n = \gamma_n - \sum_{j=1}^{n-1} \gamma_j A_{n-j}$$

The convergence is such that accurate numerical computation using the series is only feasible for $t < 2.5$ for $\beta < 4$ and decreasing to $t < 1.5$ at $\beta = 10$. When $\beta = 0.5$ then the convergence is rapid even for $t < 100$ and if $\beta = 0.8$ then reasonable time decreases to $t < 30$. Another technique proposed by [22] is based on McLaurin series. It is also not feasible since it requires n^{th} derivative of a renewal distribution. The method proposed by Constantine and Robinson [14] also for Weibull renewal process is simply to complex to employ in order to use it for only one case. It also has to be stressed that in order to compute availability by using renewal functions all methods proposed above have to be modified in order to be able to deal with delayed renewal case. Those are the main reasons why the series expansion methods are not employed later.

3.2.3 Spline approximation

Cubic spline approximation developed by McConalogue [13, 30] seems to be the first numerical method for computing renewal functions. The algorithm computes convo-

lution by integrating the spline representation of the distributions. The advantage of this approach is that it can be used directly for computing availability function and also other metrics like mission reliability. However, as mentioned by [22] this algorithm is too computationally intensive for real-time computations and it is mainly used for tabulating purposes, see [8]. Another spline approximation was proposed by Bilgen *et al* [9]. However, it is only applicable for the Weibull renewal function and it is also quite complicated. Modification for the delayed renewal process with shifted Weibull distribution is not trivial.

3.2.4 Approximations

These methods are based on weakening of certain assumptions regarding the model what leads to easier calculations. The simplest approximation to the renewal and availability functions is in the form of linear asymptotic expansion (for details see Section 2.2.1). It is of course inadequate to use such approximations on the whole interval especially when the functions have highly oscillatory behavior. Unfortunately all other methods focus on approximating a renewal function and not on an availability function. This is obviously due to the fact that availability functions require higher accuracy. As mentioned before the availability function may also be interpreted as a difference between two renewal functions. Those functions differ only by the distribution of the first renewal and therefore are very similar. Thus, the approach via renewal functions also requires high accuracy. However, such methods may be applied for approximating only renewal functions if not too much accuracy is required. Examples of such approaches may be found in [39], [25] (by approximating partial distribution); [16] (Normal approximation to Weibull renewal process); [5] (an approximate solution for renewal density).

Based on the literature review none of the above methods mentioned above is completely suitable for the problem. Some methods may only be used for renewal functions, other suffer from lack of accuracy or are too complicated to analyze. For more detailed comparison we decide to choose two left approaches: bounds and direct solution of the integral equation. This choice seems to be the best for the purpose of this research.

3.3 Bounds

3.3.1 Renewal function

A number of lower and upper bounds for a renewal function may be constructed. These bounds often depend on the assumptions about the underlying renewal distribution and may differ in shape and tightness. For example Marshall [29] presented ”the best” linear bounds for the renewal function in the ordinary renewal process as follows. If let $b_L = \inf_{t\geq 0} \frac{F(t)-F_e(t)}{1-F(t)}$ and $b_U = \sup_{t\geq 0} \frac{F(t)-F_e(t)}{1-F(t)}$, then $b_L \leq M(t) - t/\mu \leq b_U$, where μ is the expected value related to the distribution F and $F_e(t) = \frac{1}{\mu}\int_0^t (1-F(u))du$ is known as the limiting excess life distribution. Other linear bounds were obtained by Lorden [28].

Bartholomew [5] obtained the following upper bound $F(t) + \frac{1}{\mu}\int_0^t [F^2(u)/F_e(u)]du$. Other bounds may be found in e.g. [18, 24]. It has to be remarked that usually those

bounds need modification for a delayed renewal function.
For the purpose of this project it is desired to have an approximation for the renewal function not necessarily the bounds. However, if both upper and lower bounds are available then one can take for instance the mid point between the bounds as the approximation for the renewal function. Namely

$$M(t) \approx \frac{M_L(t) + M_U(t)}{2} \tag{3.5}$$

where M_L and M_U are such that $M_L(t) \leq M(t) \leq M_U(t)$, i.e. lower and upper bound for M. This approximation is appropriate only if the bounds are expected to be more or less symmetric around the unknown renewal function. In fact one could take as an approximation any point between the bounds and this would result in error

$$\varepsilon(t) = \frac{M_U(t) - M_L(t)}{2} \tag{3.6}$$

However, due to its simplicity and lack of better alternative, the midpoint approximation is used the most often. It has to be remarked that the error estimation given in (3.6) is not always useful. If the bounds are tight then the approximation (3.5) is good and the estimated error (3.6) is small. It may also happen that bounds are wide but the approximation taken as the midpoint between the bounds is good. And since there is a big difference between the bounds the error estimation obtained from (3.6) will also be large. This kind of situation will appear later in case of the availability function.
For the approximation of M_L and M_U we decide to use a simple recursive formulas that were first proposed by Ayhan *et al* [3] and later a different derivation was given by Mercier [31]. Both approaches give the same set of equations for the upper and lower bound but there were derived in a completely different manner. It is interesting that the paper of Ayhan was published earlier and Mercier does not include any comment on Ayhan's approach (it is not even in the references). Since both methods give the same algorithms we focus more deeply on just only one of them – the method of Ayhan. The method of Mercier may be briefly described as follows. Below we present the description the method of
The idea is based on constructing discrete upper and lower random bounds for the renewal times. Those bounds form a discrete renewal processes and renewal functions associated with those processes form the bounds on the continuous renewal process. Next step is to construct the renewal functions associated with those discrete renewal processes which will form the bound on the continuous renewal function.

Numerical approach: method of Ayhan

In their paper Ayhan *et al* are focused on computing bounds for a renewal function and also provide bounds on the general solution to the integral equation of the form (2.63). More details about the bounds on the solution of this integral equation will be given in Section 3.3.2. The bounds on the renewal function are exactly the same as those derived by Mercier. The derivation of the bounds is rather straightforward and is based on the direct interpretation of the Riemann-Stieltjes integral. Let us recall that the renewal function for a delayed renewal process satisfies the following integral equation

$$M_D(t) = G(t) + \int_0^t M_D(t-x)dF(x)$$

where G is the distribution of the first renewal time. Let us consider the uniform partition of the interval $[0,t]$ $0 < h < 2h < \cdots < nh = t$. Then above equation may be written as

$$M_D(t_m) = G(t_m) + \int_0^{t_m} M_D(t_m - x)dF(x) = G(t_m) + \sum_{i=1}^{m} \int_{t_{i-1}}^{t_i} M_D(t_m - x)dF(x) \quad (3.7)$$

where $m = 1,\ldots,n$. Since $M_D(t)$ is non-decreasing function and also $F(t_j) > F(t_{j-1})$ for all j the integral in (3.7) may be bounded by

$$M_D(t_m - t_i)[F(t_j) - F(t_{j-1})] \leq \int_{t_{i-1}}^{t_i} M_D(t_m - x)dF(x) \leq M_D(t_m - t_{i-1})[F(t_j) - F(t_{j-1})] \quad (3.8)$$

for all $i = 1,\ldots,m$. Therefore from (3.7) and (3.8) we obtain the lower bound for $M_D(t_m)$

$$M_D^L(t_m) = G(t_m) + \sum_{i=1}^{m} M_D^L(t_m - t_i)[F(t_j) - F(t_{j-1})] \quad (3.9)$$

and similarly the upper bound

$$M_D^U(t_m) = G(t_m) + \sum_{i=1}^{m} M_D^U(t_m - t_{i-1})[F(t_j) - F(t_{j-1})] \quad (3.10)$$

Note that right hand side in (3.9) does not depend on $M_D^L(t_m)$ thus this equations can be easily solved recursively. Right hand side in (3.10) depends on $M_D^U(t_m)$ when $i = 1$. Therefore

$$M_D^U(t_m) = G(t_m) + \sum_{i=2}^{m} M_D^U(t_m - t_{i-1})[F(t_j) - F(t_{j-1})] + M_D^U(t_m)[F(t_j) - F(t_{j-1})]$$

and solving for $M_D^U(t_m)$ gives the following recursive formula

$$M_D^U(t_m) = \frac{1}{1 - (F(t_j) - F(t_{j-1}))} \left[G(t_m) + \sum_{i=2}^{m} M_D^U(t_m - t_{i-1})[F(t_j) - F(t_{j-1})] \right] \quad (3.11)$$

Where the initial condition for both bounds is assumed to be $M_D^L(t_m) = 0 = M_D^U(t_m)$. Both authors Mercier and Ayhan suggest also matrix notation for equations (3.9) and (3.11) however we do not present it here since the recursive formulas seem quicker to compute and they are less memory size consuming.

3.3.2 Availability function

The bounds for the availability function are not so widely studied in literature as in case of the renewal function. Here we present two methods for computing the bounds for the availability function. Both methods are based on the bounds on corresponding renewal functions.

Method 1

This approach was suggested by Ayhan also in [3]. Unfortunately it is incorrect. It can be shown by mean of examples that for some cases the bounds on the availability function given in [3] are crossing. Below we present the idea and also corrected version

of the bounds. For convenience we use the same notation as in [3]. It also has to be commented that in the paper there is a small disagreement in the interpretation of the renewal process at origin. Recall that all availability functions may be found by solving the renewal type integral equation

$$k(t) = g(t) + \int_0^t k(t-s)dF(s) \tag{3.12}$$

where g, F are known and k is unknown, see also Table 2.4. Ayhan claims that the solution to (3.12) is given by

$$k(t) = \int_0^t g(t-s)dM(s) \tag{3.13}$$

where $M(t) = \sum_{j=1}^{\infty} F^{(j)}(t)$ and $F^{(j)}(t)$ is j– fold convolution of F with itself. This is statement is incorrect. Take for instance $F(t) = g(t) = 1 - e^{\lambda t}$ then $k(t) = M(t)$. It can easily be shown that

$$M(t) = \sum_{j=1}^{\infty} F^{(j)}(t) = \lambda t$$

This situation corresponds to the renewal function in Poisson process see e.g. [24]. However from (3.13) one obtains

$$k(t) = \int_0^t (1 - e^{-\lambda(t-s)})\lambda ds = \lambda t + 1 \tag{3.14}$$

This disagreement comes from the interpretation of the renewal process at origin. The statement (3.13) is true if and only if

$$M(t) = 1 + \sum_{j=1}^{\infty} F^{(j)}(t)$$

what corresponds to the situation when the first renewal is always at $t = 0$. The authors of [3] continue their notation using contradictory assumptions that (3.13) is true and $M(0) = 0$. We will keep the convention that $M(0) = 0$ and the solution to (3.12) is given by

$$k(t) = g(t) + \int_0^t g(t-s)dM(s) \tag{3.15}$$

(see also Theorem 2.2.2).
The bounds on k may be obtained in the following manner. Using the same partition of the interval the above equation may be written as

$$k(t_m) = g(t_m) + \sum_{j=1}^{m} \int_{t_{j-1}}^{t_j} g(t_m - s)dM(s) \tag{3.16}$$

Let M_L and M_U be the lower and upper bounds for M and then the integral in (3.16) can be bounded by

$$\int_{t_{j-1}}^{t_j} g(t_m - s)dM(s) \leq \max_{t_{j-1} \leq s \leq t_j} \left\{ g(t_m - s)[M(t_j) - M(t_{j-1})] \right\} \tag{3.17}$$

and

$$\int_{t_{j-1}}^{t_j} g(t_m - s)dM(s) \geq \min_{t_{j-1} \leq s \leq t_j} \left\{ g(t_m - s)[M(t_j) - M(t_{j-1})] \right\} \tag{3.18}$$

Of course M is unknown and the bounds for it have to be used instead. The authors in [3] suggested to use the following lower bound

$$\min_{t_{j-1} \leq s \leq t_j} \left\{ g(t_m - s)[M_L(t_j) - M_L(t_{j-1})] \right\} \leq \min_{t_{j-1} \leq s \leq t_j} \left\{ g(t_m - s)[M(t_j) - M(t_{j-1})] \right\}$$

Which is based on their Proposition 3 in which their prove that

$$M_L(t_j) - M_L(t_{j-1}) \leq M(t_j) - M(t_{j-1}) \tag{3.19}$$

Upper bound is constructed in similar way. It is not that difficult to find an example for which (3.19) does not hold. Let us consider Gamma renewal process (this process is introduced in Section 3.5.1) for which analytical expressions for M is known. This process gives a flexibility to model either oscillating and non-oscillating renewal functions. It appears that inequality (3.19) holds only when the oscillations are small (such examples, only with small oscillations, were presented in [3]). In Figure 3.1 we present renewal functions for two cases with oscillatory and non-oscillatory behavior. The renewal function M was obtained from an explicit formula given later in Section 3.5.1, bounds were calculated using the recursive formulas (3.9) and (3.11). From this

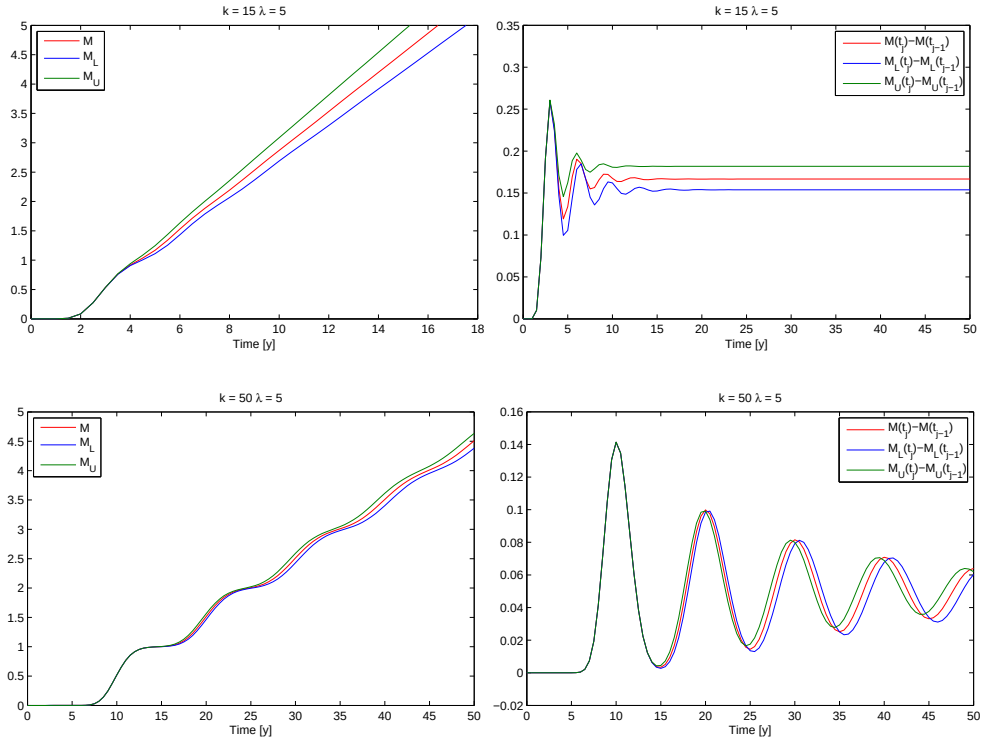


Figure 3.1: Bound on the renewal function for Gamma renewal process and corresponding increments of the bounds. Number of grid points used $N = 100$ for both cases.

Figure it may be observed that both bounds are coherent with the analytical solution. Plots on the right hand side present the increments for both bounds and also for the renewal function. In case when the oscillations are small the increments of the bounds bound the increment of the renewal function. When the oscillations become bigger then this is not true. Since in such cases relation (3.19) does not hold the bounds on k given in [3] will also be crossing.
To fix this problem we propose the following. Since $M_L(t)$ and $M_U(t)$ form bounds on $M(t)$ for every $j = 1,2,\ldots$ it holds

$$M_L(t_j) - M_U(t_{j-1}) \leq M(t_j) - M(t_{j-1}) \leq M_U(t_j) - M_L(t_{j-1}) \tag{3.20}$$

Therefore the integral in (3.16) may be bounded by

$$\int_{t_{j-1}}^{t_j} g(t_m - s)dM(s) \leq \max_{t_{j-1}\leq s\leq t_j} \left\{g(t_m - s)[M_U(t_j) - M_L(t_{j-1})]\right\} \tag{3.21}$$

and

$$\int_{t_{j-1}}^{t_j} g(t_m - s)dM(s) \geq \min_{t_{j-1}\leq s\leq t_j} \left\{g(t_m - s)[M_L(t_j) - M_U(t_{j-1})]\right\} \tag{3.22}$$

Note that since M is non-decreasing the upper bound is bigger than the lower bound for all (different) arguments, hence $M_U(t_j) - M_L(t_{j-1}) \geq 0$. Therefore the upper bound on the function k is

$$k(t_m) \leq g(t_m) + \sum_{j=1}^{m} \max_{t_{j-1}\leq s\leq t_j} \{g(t_m - s)\} [M_U(t_j) - M_L(t_{j-1})] \tag{3.23}$$

This bound may be easily computed if the function $g(t)$ is monotone. If $g(t)$ is increasing then $\max_{t_{j-1}\leq s\leq t_j} \{g(t_m - s)\} = g(t_m - t_{j-1})$ and if $g(t)$ is decreasing then $\max_{t_{j-1}\leq s\leq t_j} \{g(t_m - s)\} = g(t_m - t_j)$. From (3.22) and (3.16) it follows that the lower bound on k is

$$k(t_m) \geq g(t_m) + \sum_{j=1}^{m} \min_{t_{j-1}\leq s\leq t_j} \left\{g(t_m - s)[M_L(t_j) - M_U(t_{j-1})]\right\} \tag{3.24}$$

This bound may also be computed very easily when $g(t)$ is monotone however one have to be careful about the sign of $M_L(t_j) - M_U(t_{j-1})$. Therefore for the above minimum one should take $g(t_m - t_j)$ if $g(t)$ is increasing (decreasing) and $M_L(t_j) - M_U(t_{j-1}) > 0$ ($M_L(t_j) - M_U(t_{j-1}) < 0$). And $g(t_m - t_{j-1})$ if $g(t)$ is increasing (decreasing) and $M_L(t_j) - M_U(t_{j-1}) < 0$ ($M_L(t_j) - M_U(t_{j-1}) > 0$).
In the example below we check the behavior of these bounds in case of the availability function A_1.

Example 3.3.1. *Let up times U_i be distributed according to Weibull distribution with parameters α and β and let the down-times D_i be constant. Then the distribution F_{U+D} is a shifted Weibull distribution and is given by (3.56). We are interested in the availability function $A_1(t)$ for different parameters. This corresponds to $k(t) = A_1(t)$, $F(t) = F_{U+D}$ and $g(t) = 1 - F_U(t)$. Figure 3.2 presents the bounds obtained from formulas (3.23) and (3.24) and also the approximation taken as the mid point between*

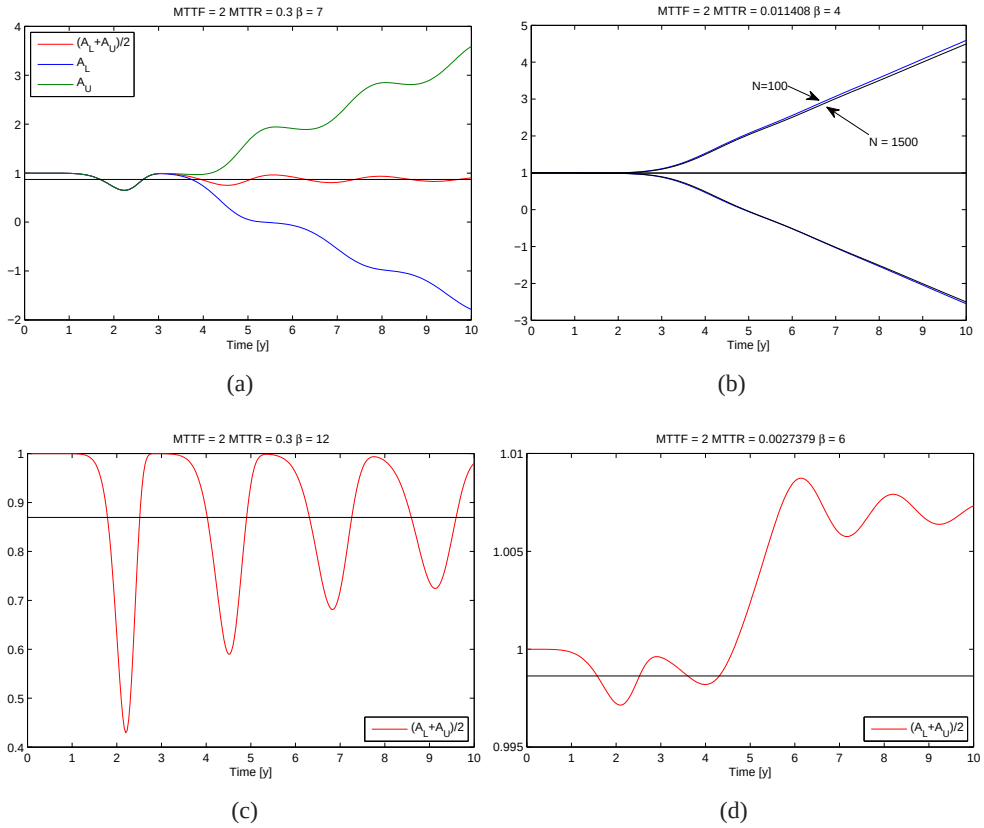


Figure 3.2: Behavior of the bounds on the availability function A_1 for Weibull-Constant case.

bounds. The first observation from the plot 3.2(a) is that the bounds on the availability constructed from the bounds on the renewal function are very wide for bigger values of t*. However, it is interesting that although bounds are wide the approximation taken as the midpoint converges to the steady state and it also reflects oscillatory behavior. Graph 3.2(c) presents only the approximation obtained from the bounds. Since the bounds on the availability are very wide (very often beyond the interval* $[0,1]$*) it is natural to ask how fast the bounds will become tighter if the number of grid points will be increased. The answer is given on the graph 3.2(b) which presents the lower and upper bounds on the availability function for two cases: one with 100 grid points and second one with 1500 points. It can be seen that increasing the number of the grid points by 15 times leads to a very small change in the bounds. In general this is the problem with discretization methods that they have limited accuracy. Finally let us focus on more realistic values of the parameters. Plot 3.2(d) illustrates the behavior of the the approximation for the availability function obtained from the bounds when the steady state is equal to 0.9986. It can be observed that approximated availability is above one (for this case we used 1500 points). This is the general behavior for high availability states. The conclusion from this example is as follows: this method is reasonable for the cases with low steady states (*< 0.9*). For the situations when steady state is high this method gives a poor result.*

Method 2

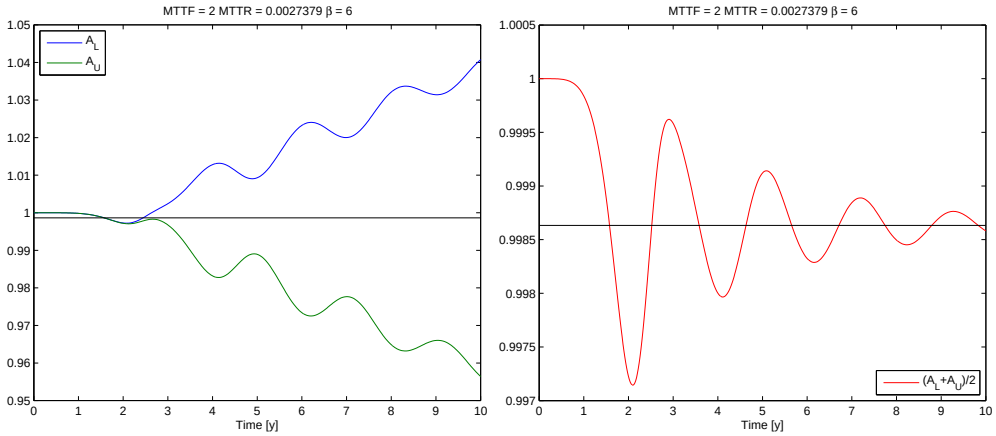
This method bounds the availability function also using bounds on the renewal function. The advantage of this method is that it can also handle the availability functions with high steady state. The idea is very simple and is based on the relation between the availability function and two renewal functions. Let us present the method in case of function A_1. Recall that the availability function A_1 may be obtained from

$$A_1(t) = M_{1R}(t) - M_{1F}(t) + 1$$

for other availability functions see Section 2.3. Let $M^L_{1R}(t)$ and $M^U_{1R}(t)$ be the bounds for the function $M_{1R}(t)$, similarly let $M^L_{1F}(t)$ and $M^U_{1F}(t)$ denote the bounds for $M_{1R}(t)$ then from above relation it follows that

$$M^L_{1R}(t) - M^U_{1F}(t) + 1 \leq A_1(t) \leq M^U_{1R}(t) - M^L_{1F}(t) + 1 \tag{3.25}$$

Therefore we obtain the bounds on the availability function. In principle one can use any bounds on the renewal function and use relation (3.25) to obtain bounds on the availability function. Of course the drawback of this method is that it requires computing four bounds on two different renewal functions. On the other hand once such bounds are computed the bounds on the availability are computed in a negligible time. Let us consider again the case presented on Figure 3.2(d). It was remarked in the example above that the Method 1 had problems with the high availabilities. It appears that Method 2 can handle high availability states much more efficiently. Figure 3.3 presents the bounds on the availability and the approximation (separately on two plots) for the same case as the one presented on Figure 3.2(d). Bounds on the renewal functions were computed by using the same recursive formulas. For this computation we used only 500 grid points. Note that the bounds on the availability function are also



(a) Bounds on the availability function. (b) The approximation for the availability function

Figure 3.3: Approximation to the renewal function for the case taken from the Example 3.3.1

quite wide and their usage as the error estimation is rather doubtful. The approximation (mid point between the bounds) presented on Figure 3.3(b) suggests that this method may be considered as a good candidate for approximating the availability functions.

Remark 3.3.2. Run *et al* in [38] present a method for improving any bound for a renewal function by a functional iteration. This procedure may be described as follows. If $M_L(t)$ and $M_U(t)$ form the bounds on the renewal function $M(t)$ then those bounds may be used as the initial conditions in the following iterative scheme

$$M_L^{(i+1)}(t) = F(t) + \int_0^t M_L^{(i)}(t-s)dF(s)$$

or

$$M_U^{(i+1)}(t) = F(t) + \int_0^t M_U^{(i)}(t-s)dF(s)$$

It is proven that $M_L^{(i)}$ and $M_U^{(i)}$ are bounds for all i and converge to the renewal function when i increases. Although this iterative procedure may be used to improve the bounds obtained in this section we do not employ it because it is computationally inefficient. Each iteration leads to the integration of the renewal function obtained from the previous iteration therefore the initial bound has to be computed on a very fine grid. This drawback can also be observed in their paper, namely all values of the improved bounds are presented only at few grid points. This method may be used only if the the single values of the renewal function are required.

3.4 Numerical solution of the integral equation

In previous Chapter it was shown that all functions of interest satisfy the same type of integral equation

$$r(t) = H(t) + \int_0^t r(t-x)dF_{U+D}(x) \tag{3.26}$$

where H, F_{U+D} are given and r is unknown. Although we say that F_{U+D} is given in most of the cases it has to be approximated as well. However it is more convenient to consider the problems of computing F_{U+D} and r separately. An obvious numerical procedure for solving (3.26) is to approximate the integral term via quadrature rule which integrates over the variable x for a fixed value of t. Below we present three quadratures that appear to be the most suitable for the problem.
Assume that the solution to (3.26) is required on the interval $[0,t]$. Let $0 = t_0 < t_1 < \cdots < t_n < t$ be a partition of the interval. As it was already seen in Section 3.3 the functions r may have oscillatory behavior, like for instance the availability functions. One may think that choosing a non-uniform partition of $[0,t]$ would result in higher accuracy. Unfortunately the non-uniform grid increases the required number of calculations of F_{U+D}. We will discuss this problem in more details in Section 3.4.4 where we also show that not big improvement is achieved when the unequal spacing is used. Therefore we choose to use the uniform grid $t_i = ih, i = 1, \ldots, n$ $h > 0$. The equation (3.26) may then be written as

$$r(t_m) = H(t_m) + \int_0^{t_m} r(t_m - x)dF_{U+D}(x) = H(t_m) + \int_0^{t_m} r(t_m - x)dF_{U+D}(x) \tag{3.27}$$

3.4.1 Trapezoid and Simpson's rules

Tortorella in [47] presented two approximations for $r(t_i)$ based on the trapezoid and Simpson's rule for Riemann-Stieltjes integrals. The trapezoid rule can be derived in the

same manner as for Riemann integrals by replacing the increment dx by $dF_{U+D}(x)$. For methods with orders 2 (Simpson's rule) Tortorella proposes new Newton-Cotes rule for Riemann-Stieltjes integrals, see also [46]. Direct replacement of dx by $dF_{U+D}(x)$ for higher order methods will result in accuracies that are not as good as those of the rules obtained in [47]. The derivation of the recursive formulas for approximating $r(t_i)$ may be found in [47]. Here we present the end-equations and stress the important aspects related to our problem. For convenience we use the same notation as in [47]. Let F_i denote the value of $F_{U+D}(t_i)$ and similarly for $H_i = H(t_i)$, $r_i = r(t_i)$. The algorithms are based on the assumption that $F(0) = 0$ and $H(0) = 0$ (whence $r(0) = 0$). Seemingly this restricts the usage of the algorithms to only renewal functions. The algorithms may also be used for the case when $H(0) \neq 0$ with simple scaling procedure that is presented in Remark 4.2.3. Below are the algorithms

Trapezoid Rule

$$r_i = \frac{2H_i}{2-F_1} + \sum_{k=1}^{i-1} r_{i-k} \frac{F_{k+1} - F_{k-1}}{2-F_1} \tag{3.28}$$

This procedure requires one starting point: $r_0 = 0$.

Simpson's Rule This procedure requires three starting points: $r_0 = 0$, $r_{1/2}$ and r_1, where $r_{1/2} = r(t_1/2)$. In principle those values could be obtained using trapezoid rule, however the error introduced in initial points has significant influence on the whole solution (see [47]). In order to find those initial points Tortorella propose so called ascending-descending procedure. Let $s_k = 2^{-k} t_1$, $k = 0, 1, 2, \ldots$. Note that sequence s_k descends to zero as k increases and also that $s_0 = t_1$, $s_1 = t_1/2$. This grid is presented on Figure 3.4. The idea is to take sufficiently large K such that s_{K+1} is almost zero and use Simpson's rule with starting values $r_0 = 0$ and $r(s_K + 1) = 0$ to obtain an estimate for $r(s_K)$ and 'descend' back up to $r(t_1)$. This ascending-descending procedure can be

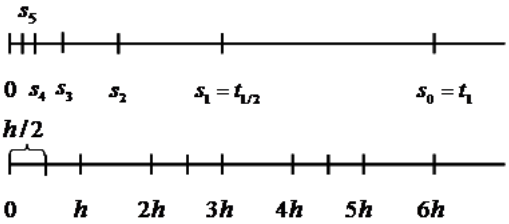


Figure 3.4: Grid for ascending-descending procedure (upper) and for Simpson's rule (lower)

described as follows. Choose sufficiently large K (see Remark 3.4.2). Compute

$$r(s_K) = (1 + F(s_K))/6)^{-1} H(s_K)$$

Take $i = 1, \ldots, K$ and for $k = K - 1$ compute

$$r(s_k) = \left[1 - \frac{2F(s_{k+1})}{3} + \frac{F(s_k)}{6}\right]^{-1} \left[H(s_k) + \frac{2F(s_k)}{3} r(s_{k+1})\right]$$

Then the starting values for the Simpson's rule are $r_{1/2} = r(s_1)$, $r_1 = r(s_0)$.
The equations for approximating r_i by using Simpson's rule derived in [47] are

$$\begin{aligned} r_i &= \left(1 - \frac{2F_1}{3} + \frac{F_2}{6}\right)^{-1} \left[H_i + \frac{2}{3}F_2 r_{i-1} + r_{i-2}\left(\frac{1}{2}F_2 - \frac{2}{3}F_1\right) + \right. \\ &+ \sum_{k=2}^{i/2} \left[\frac{r_{i-2k} + 4r_{i-2k+1} + r_{i-2k+2}}{6}(F_{2k} - F_{2k-2}) + \right. \\ &+ \left.\left. \frac{r_{i-2k} - r_{i-2k+2}}{3}(F_{2k} - 2F_{2k-1} + F_{2k-2})\right]\right] \end{aligned}$$

for $i = 2, 4, \ldots$, and

$$\begin{aligned} r_i &= \left(1 - \frac{2F_1}{3} + \frac{F_2}{6}\right)^{-1} \left[H_i + \frac{2}{3}F_2 r_{i-1} + r_{i-2}\left(\frac{1}{2}F_2 - \frac{2}{3}F_1\right) + \right. \\ &+ \sum_{k=2}^{(i-1)/2} \left[\frac{r_{i-2k} + 4r_{i-2k+1} + r_{i-2k+2}}{6}(F_{2k} - F_{2k-2}) + \right. \\ &+ \left. \frac{r_{i-2k} - r_{i-2k+2}}{3}(F_{2k} - 2F_{2k-1} + F_{2k-2})\right] \\ &+ \left. \frac{4r_{1/2} + r_1}{6}(F_i - F_{i-1}) - \frac{r_1}{3}(F_{i-1} - 2F_{i-1/2} + F_i)\right] \end{aligned}$$

for $i = 3, 5, \ldots$.

Remark 3.4.1. The algorithms presented in this section approximate function r given that $r(0) = 0$. In order to use them for general case for which $r(0) \neq 0$ it is needed to scale the function H. Let $H^z(t) = H(t) - H(0)(1 - F(t))$ and $r^z(t) = r(t) - H(0)$ then it is easy to find that (r^z, H^z, F) satisfy (3.26) if and only if (r^z, H^z, F) satisfy (3.26). So for instance when computing the availability function A_1, see equation for which $A_1(0) = 1$ we need to first compute

$$H^z(t) = H(t) - 1 - F(t)$$

where $H(t) = 1 - F_U(t)$ and $F(t) = F_{U+D}(t)$. Then using H^z as a proper input to the algorithms one can compute values of $r^z(t)$. After the approximation $r^z(t)$ is obtained it has to be re-scaled in order to have the solution for $r(t)$ ($A_1(t)$). This is done using relation mentioned above

$$r(t) = r^z(t) + 1.$$

It has to be stressed that this scaling procedure is done in a negligible time.

Remark 3.4.2. *Choice of K in ascending-descending procedure.* The choice of K in the ascending-descending procedure will determine the error in the starting points $r(t_1)$ and $r(t_{1/2})$. It appears that the error introduced in those initial points may have significant impact on the overall approximation. However the ascending-descending procedure eliminates the error for the initial conditions very efficiently. Based on the experiments we claim that for $K \geq 7$ the difference between two approximations with different K is beyond the machine precision. We could not find any example for which it is not true. The cost of computing 7 approximations is negligible.

The very big advantage of the Simpson's rule derived by Tortorella is that it requires the values of the distribution F_{U+D} at only $N+N/2+K$ grid points (see Figure 3.4). Usually the second order methods require $2N$ values for the function F_{U+D}.

3.4.2 Midpoint Rule

This method is very similar to the trapezoid one but instead it uses the mid point values for the function F_{U+D}. Below we present the derivation. Based on Remark 2.1.7 the equation (3.26) may be written as

$$r(t_m) = H(t_m) + \int_0^{t_m} r(t_m - x) dF_{U+D}(x) \tag{3.29}$$

$$= H(t_m) + \int_0^{t_m} F_{U+D}(t_m - x) dr(x) + F(t) r(0) \tag{3.30}$$

The integral in the last line ma be approximated by the midpoint rule

$$\int_0^{t_m} F_{U+D}(t_m - x) dr(x) = \sum_{j=1}^{m} \int_{t_{j-1}}^{t_j} F_{U+D}(t_m - x) dr(x)$$

$$\approx \sum_{j=1}^{m} F_{U+D}(t_m - t_{j-0.5})[r(t_j) - r(t_{j-1})] \tag{3.31}$$

where $t_{j-0.5} = (t_j - t_{j-1})/2$. Substituting in (3.29) and solving for $r(t_m)$ gives

$$r(t_m) = \frac{1}{1 - F(t_{0.5})} \left[H(t_m) + r(0) F_{U+D}(t_m) - r(t_{m-1}) F_{U+D}(t_{0.5}) + \sum_{j=1}^{m-1} F_{U+D}(t_m - t_{j-0.5})[r(t_j) - r(t_{j-1})] \right] \tag{3.32}$$

The disadvantage of this recursion formula is that it requires the values of the distribution F_{U+D} on the grid with the step size $h/2$. This is an important issue when F_{U+D} has to be approximated.

3.4.3 Volterra equations

If F_{U+D} has a density f_{U+D} then (3.26) may be written as

$$r(t) = H(t) + \int_0^t r(x) f_{U+D}(t - x) dx \tag{3.33}$$

Equation (3.33) is a linear Volterra integral equation of the second kind with a difference kernel. Numerical solution of this type of equation has become routine, see for example [17]. However, interest remains in dealing with (3.26) directly because:

- We may compute all functions of interest (see Table 2.4) by only using F_{U+D} without involving the density f_{U+D}. Some functions like A_1 or A_0^D may be computed using only density f_{U+D} however in order to compute functions A_0 or A_1^D one has to compute F_{U+D} either by integrating the density (computed for A_0 or A_1^D) or by using the algorithm for convolution. This double effort when computing more functions is eliminated when using form (3.26).
- We may want to avoid problems when f_{U+D} has singularity.

- As mentioned in [47], [19] simpler, more robust or more accurate methods may arise from a direct approach (3.26)

Nevertheless we do not reject this approach immediately. Since the literature for solving numerically Volterra integral equations is rather rich we decide to choose one representative method from this class and to compare its performance for computing the renewal and availability functions with other methods. The available methods for solving Volterra integral equations are: quadrature methods, Runge-Kutta methods, block methods just to mention few. For the overview of the available methods see e.g. [17]. As the representative of this class we decide to choose fourth order Runge-Kutta method (RK-4). The main reason is that it differs significantly from the methods presented so far and also because the order of this method is higher. Runge-Kutta methods for the solution of (3.33) are self-starting methods which determine approximations to the solution at the points $t_i = hi$, $i = 1,2,\ldots$ by generating approximations at some intermediate points in $[t_i, t_{i+1}]$ (RK-4 uses three points). The derivation of the algorithm follows closely that for ordinary differential equations and may be found for example in [17]. The algorithm itself is not difficult but slightly complex and is not presented here for typographical reasons (see p. 123 in [17]).
We will check the performance of RK-4 method together with other methods in Section 3.5.

3.4.4 Non-uniform grid

The methods described in previous sections are fixed step methods, that is, they are methods which divide the range of interest $[0,t]$ into N equally spaced intervals of length $h = t/N$ and which solve the integral equation at the set of discrete points $t_i = ih$. Since the availability and renewal functions tend to have an oscillatory shape it is also important to investigate the behavior of the approximation on a non-uniform grid. It may happen that putting more points in the places where the approximated function is changing more rapidly gives better approximation than in the case of equal spacing.
In this section we check how the change of the grid influences the approximation and what is the actual cost of using non-uniform grids. The results are presented for two methods: trapezoid rule and method with bounds. In order to use the methods proposed by [47] and [31, 3] on the non-uniform grids the modification of the algorithms is necessary. For the trapezoid rule it is not difficult to show that the formula for approximating $r(t_i)$ below may be used on any partition $0 = t_0 < t_1 < \cdots < t_N = t$

$$r(t_i) = \frac{2}{2 - F(t_i - t_{i-1})}\left[H(t_i) + \frac{r(t_{i-1})}{2}F(t_i - t_{i-1}) + \sum_{k=1}^{i-1}\frac{r(t_k) + r(t_{k-1})}{2}\left[F(t_i - t_{k-1}) - F(t_i - t_k)\right]\right] \tag{3.34}$$

For the method with bounds we present the results only for the renewal function. The modified recursive formulas for the bounds are

$$M_L(t_m) = F(t_m) - \sum_{i=1}^{m} M_L(t_{i-1})\left[F(t_m - t_i) - F(t_m - t_{i-1})\right] \tag{3.35}$$

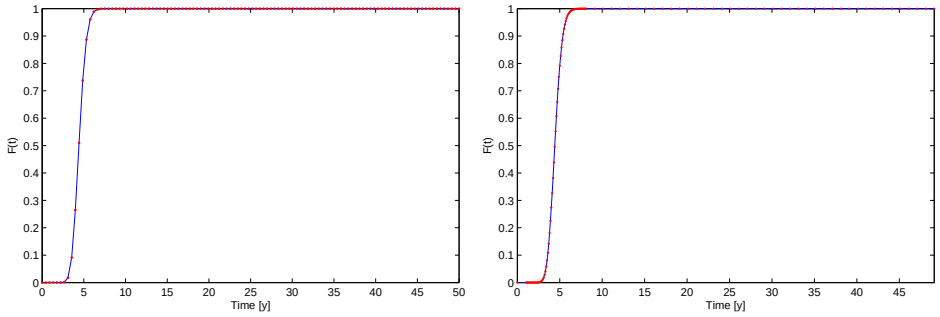
$$= F(t_m) + \sum_{i=1}^{m} M_L(t_{i-1})\left[F(t_m - t_{i-1}) - F(t_m - t_i)\right] \tag{3.36}$$

and

$$M_U(t_m) \quad = \quad F(t_m) + \sum_{i=1}^{m} M_U(t_i)\left[F(t_m - t_{i-1}) - F(t_m - t_i)\right]$$

where we used the notation $F(t) = F_{U+D}(t)$. Those bounds may be used on any partition.
Note that if the grid is uniform then $t_k - t_i = t_{k-i}$ and the number of function evaluations of F is minimal and equal to N. If the non-uniform grid is used then the number of function evaluations needed will increase and it will depend on the structure of the grid (in the worse case it could be $(N^2 - N)/2$). Since in many cases F has to be approximated it is important to check if the usage of non-uniform grid will compensate this additional cost.
The general idea behind the non-uniform grid can be described as follows. Having N points in the grid find best possible location for those points in the sense that the approximation obtained on this grid has the smallest overall error. Of course it is not obvious where is the best location, however we investigate it experimentally. In the example below we compare the approximations obtained from the trapezoid rule on uniform and non-uniform grids.

Example 3.4.3. *Assume that the distributions of up times and down times have both Gamma distribution with parameters: $U_i \sim Gamma(39,9)$ and $D_i \sim Gamma(1,9)$ [2], then $F(t) = F_{U+D}(t) \sim Gamma(40,1)$ (for details about Gamma alternating renewal process see Section 3.5.1). Our interest is in the availability function A_1. Let us say that we have $N = 114$ points at our disposal. From those points we decide to construct two grids: (1) uniform and (2) non-uniform. The best distribution of the points on the non-uniform grid is not obvious; we decide to put more points in place when the distribution F is changing rapidly (this is also the place where the function A_1 is changing rapidly). The two grids are presented in Figure 3.5. The approximations for the availability*



(a) Uniform grid. (b) Finer grid where F is changing.

Figure 3.5: Two grids from Example 3.4.3

function obtained from the trapezoid rule on those two grids are shown in Figure 3.6. From this graph it can be observed that the overall approximation obtained on the non-uniform grid is much worse than the one in the uniform case.

[2] Other distributions could be taken as well, however for Gamma case the analytical solution is known

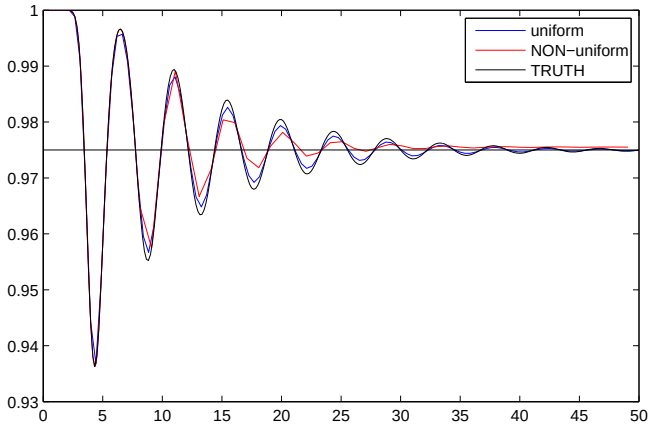
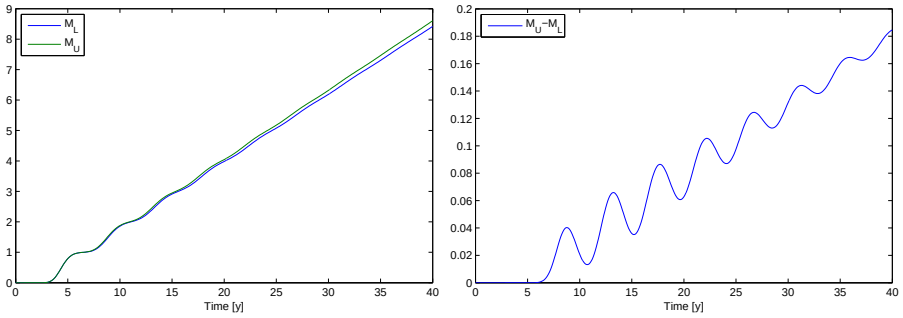


Figure 3.6: Comparison of the approximations for the availability function on different grids from Example 3.4.3

In the next example we investigate the behavior of the bounds on the renewal function using a non-uniform grid. Since the bounds on the renewal function determine the bounds on the availability function and those are used to compute the approximation for the availability function we decide to investigate only the behavior of the bounds on the renewal function.

Example 3.4.4. *Let us consider a renewal process with the underlying distribution* $F \sim Gamma(40,9)$. *Define the following uniform grid:* $t^1 = [0(0.1)40]$. *Figure 3.7(a) illustrates the bounds on the renewal function computed on this grid. The decision*



(a) Bounds on the renewal function. (b) Difference between the bounds.

Figure 3.7: Bounds on the renewal function from Example 3.4.4 computed on the uniform grid.

on how to construct a non-uniform grid is based on the distance between the bounds (see Figure 3.7(b)). Note that the difference between the bounds is bigger when the bounds are increasing. It seems to be the natural choice to put more points in the place where the distance between the bounds is relatively big. Therefore we construct the following non-uniform grid $t^2 = [0(0.05)10,\ 10.1(0.15)40]$, *which has the same number of points as grid* t^1 *(*$N = 401$*) and puts more points between* $[5,10]$. *Figure 3.8 presents the difference between upper and lower bound computed on those two grids:*

t^1 and t^2. The bounds obtained on the non-uniform grid are tighter on the interval

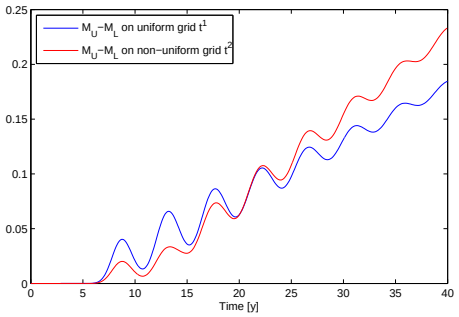


Figure 3.8: Comparison of the distance between the bounds on different grids from Example 3.4.4.

$[0, 20]$. *For $t > 20$ the bounds computed on the uniform grid are better.*

Based on Examples 3.4.3 and 3.4.4 we can conclude that

1. The non-uniform distribution of the points in the grid does not improve the overall approximation for the availability function.
2. Bounds for the renewal function can be tighter on the non-uniform grid but only on bounded interval.
3. It is not worth to use non-uniform grids since the benefit is small and the cost attained by the additional computation of the renewal distribution is big.

Therefore in the rest of computations we use uniform grid for computing renewal and availability functions.

3.5 Comparison of the methods

In the previous section we have introduced different methods for approximating renewal and availability functions. The aim of this section is to illustrate the comparative performance of those methods by using them to solve a series of test problems and to determine which routine is 'best' in the sense of timing and accuracy. Of course, to define what we mean by 'best' we need to declare a particular test strategy which can be applied uniformly to each routine. Inconvenience in our situation is the lack of test cases for which the analytical solution is known, especially for the availability functions. The exponential case, for which the solution was given in Chapter 2, is not suitable for comparing the overall performance of the methods. Although, it may be used as an easy test case. As it was mentioned in the requirements we are particularly interested in situations where the availability and renewal functions show oscillatory behavior as it is for the case with Weibull uptime for higher values of the shape parameter. In the literature there are a few examples available for which the renewal function is slightly oscillating and the analytical solution is known. The availability function A_1 for the case when up-times are exponentially distributed and down-times are constant was presented in [4, 24]. Unfortunately, in this situation the oscillations are visible

only for low steady states, thus not of our interest.
In the absence of adequate test cases with known solutions not much can be done with respect to the comparison of the methods. Fortunately, it appears that it is possible to derive explicit formulas for the availability and renewal functions in case when the up-times and downtimes have both Gamma distribution with an integer shape parameter. In this situation the scale and the amplitude of the oscillations may be easily controled what allows us to compare the methods for the problems with different difficulty level. The availability and renewal functions will be of course different than in case when the up-time is Weibull but the general behavior will be similar. Therefore we compare the methods mainly based on their performance for Gamma alternating process and we claim that the method selected based on this choice will also be able to produce reasonable results for the case when the up-time has Weibull distribution.
Before the comparison of the methods the derivation of the availability and renewal functions for the Gamma renewal alternating process is given.

3.5.1 Availability in Gamma alternating renewal process

Availability function for Gamma alternating renewal process was previously studied by Pham-Gia and Turkkan in [36]. They propose to use Gamma distribution for up time $U_i \sim Gamma(\alpha_U, \beta_U)$ and down time $D_i \sim Gamma(\alpha_D, \beta_D)$ where $\alpha_U, \alpha_U, \beta_U, \beta_D \in \mathbb{R}_+$(for parametrization of the Gamma distribution see Appendix A). Their method can be described as follows: compute the density $f_{U+D}(t)$ of $U_i + D_i$ (which has to be approximated by using, rather complicated, Humbert function), next approximate renewal density $m(t)$ by iterative equation

$$m^{(i)}(t) = f_{U+D}(t) + \int_0^t m^{(i-1)}(t)\, f_{U+D}(x)dx$$

with starting point taken from [5]

$$m^{(0)}(t) = f_{U+D}(t) + F^2_{U+D}(t) \Big/ \int_0^t \left(1 - F^2_{U+D}(t)\right) dx$$

The availability function A_1 is then obtained by numerical integration using relation (2.33). This method is rather complex for real-time computations and we do not employ it. Instead we propose to use Gamma distribution with the same shape parameter for up time and down time and integer scale parameter for which closed formulas for availability and renewal functions can be obtained. When scale parameter in Gamma distribution is integer then this distribution is called Erlang distribution. Hwank [45] give a numerical solution for the availability function when U_i and $U_i + D_i$ have both Erlang distribution. We present analytical expressions for the renewal and availability functions that are easy to implement and are exact.
Let $U_i \sim Gamma(k-m, \lambda)$ and $D_i \sim Gamma(m, \lambda)$ where $m < k$ are integers and $\lambda > 0$. Then $U_i + D_i \sim Gamma(k, \lambda)$. The up-times can be interpreted as a sum of $k-m$ exponentially distributed random variables with parameter λ and similarly down-times correspond to a sum of m random variables with $\sim Exp(\lambda)$. Recall that the availability function A_1 can be expressed as a difference of two renewal functions (see Section 2.4)

$$A_1(t) = M_{1R}(t) - M_{1F}(t) + 1 \tag{3.37}$$

where $M_{1R}(t)$ denotes expected number completed repairs and denotes expected number $M_{1F}(t)$ failures in $(0,t]$(see Figure 2.1). We first compute the renewal functions and then the availability is obtained from the above relation.
The main result is based on the following theorem

Theorem 3.5.1. *For any real number u and positive integers n, k, m such that $k \geq m$ it holds*

$$\sum_{n=1}^{\infty} \frac{u^{nk-m}}{(nk-m)!} = \frac{1}{k} \sum_{r=0}^{k-1} \varepsilon^{mr} e^{u\varepsilon^r} \tag{3.38}$$

where $\varepsilon = \exp\left(\frac{2\pi i}{k}\right)$, $\varepsilon^0 = 1$, $\varepsilon^r = \exp\left(\frac{2\pi r i}{k}\right)$ *and i is imaginary unit satisfying* $i^2 = -1$.

For the proof see Appendix D. This theorem expresses infinite series in terms of finite sum that involve complex numbers. Such a representation allows to simplify the calculations and is crucial in obtaining the analytical expressions.
We start with computing M_{1R}. Since there is no closed form for Gamma CDF it is easier to derive expressions for corresponding renewal densities and use relation (2.2) to get the renewal functions.

$$M_{1R}(t) = \int_0^t m_{1R}(x)dx = \int_0^t \sum_{n=1}^{\infty} f_{U+D}^{(n)}(x)dx \tag{3.39}$$

Since the sum of Gamma distributed random variables has Gamma distribution with the same shape parameter and sum of the scale parameters the convolution of $f_{U+D}^{(n)}(t)$ is $Gamma(nk,\lambda)$ i.e.

$$f_{U+D}^{(n)}(t) = \frac{\lambda}{(nk-1)!}(\lambda t)^{nk-1} e^{-\lambda t}$$

Therefore

$$M_{1R}(t) = \int_0^t \lambda e^{-\lambda x} \sum_{n=1}^{\infty} \frac{(\lambda x)^{nk-1}}{(nk-1)!} dx \tag{3.40}$$

By Theorem 3.5.1 above series can be written as

$$\begin{aligned} M_{1R}(t) &= \int_0^t \lambda e^{-\lambda x} \frac{1}{k} \sum_{r=0}^{k-1} \varepsilon^r e^{\lambda x \varepsilon^r} dx \\ &= \int_0^t \lambda e^{-\lambda x} \frac{1}{k} \left(e^{\lambda x} + \sum_{r=1}^{k-1} \varepsilon^r e^{\lambda x \varepsilon^r} \right) dx \\ &= \frac{\lambda t}{k} + \frac{1}{k} \sum_{r=1}^{k-1} \lambda \varepsilon^r \int_0^t e^{-\lambda x(1-\varepsilon^r)} dx \end{aligned}$$

By doing simple integration it can be shown that the renewal function M_{1R} is

$$M_{1R}(t) = \frac{\lambda t}{k} + \frac{1}{k} \sum_{r=1}^{k-1} \frac{\varepsilon^r}{1-\varepsilon^r} \left(1 - e^{-\lambda t(1-\varepsilon^r)}\right) \tag{3.41}$$

The same formula was obtained by Parzen [35]. However, the result presented here is more general and also it allows to calculate the availability functions which was not

given in [35]. Now we present similar approach for computing M_{1F}. Recall that in this case inter arrivals are defined as $U_1, D_1+U_2, D_2+U_3\ldots$. This situation corresponds to the delayed renewal process with first interval distribution $Gamma(k-m,\lambda)$ and with remaining arrivals distributed according to $Gamma(k,\lambda)$ (for details about this renewal process see Section 2.4). Therefore by applying the same argument as for M_{1R} it can be shown that

$$\begin{aligned} M_{1F}(t) &= \int_0^t \sum_{n=1}^{\infty} \frac{\lambda}{(nk-m-1)!}(\lambda x)^{nk-m-1}e^{-\lambda x}dx \\ &= \int_0^t \lambda e^{-\lambda x}\sum_{n=1}^{\infty}\frac{(\lambda x)^{nk-(m+1)}}{(nk-(m+1))!}dx \end{aligned}$$

By Theorem 3.5.1

$$\begin{aligned} M_{1F}(t) &= \int_0^t \lambda e^{-\lambda x}\frac{1}{k}\sum_{r=0}^{k-1}\varepsilon^{r(m+1)}e^{\lambda x\varepsilon^r}dx \\ &= \frac{\lambda t}{k}+\frac{1}{k}\sum_{r=1}^{k-1}\lambda\varepsilon^{r(m+1)}\int_0^t e^{-\lambda x(1-\varepsilon^r)}dx \end{aligned}$$

Therefore the renewal function M_{1F} can be expressed as

$$M_{1F}(t) = \frac{\lambda t}{k}+\frac{1}{k}\sum_{r=1}^{k-1}\frac{\varepsilon^{r(m+1)}}{1-\varepsilon^r}\left(1-e^{-\lambda t(1-\varepsilon^r)}\right) \tag{3.42}$$

Substituting (3.41) and (3.42) in (3.37) we obtain that the availability function A_1 for Gamma alternating process can be expressed as

$$A_1(t) = \frac{1}{k}\sum_{r=1}^{k-1}\frac{\varepsilon^r}{1-\varepsilon^r}\left(1-e^{-\lambda t(1-\varepsilon^r)}\right)(1-(\varepsilon^r)^m)+1 \tag{3.43}$$

In the same manner one can obtain availability function A_0. Let as before $U_i \sim Gamma(k-m,\lambda)$ and $D_i \sim Gamma(m,\lambda)$. Recall that the availability function A_0 may be found by

$$A_0(t) = M_{0R}(t) - M_{0F}(t) \tag{3.44}$$

As it was mentioned in Section 2.3.3 the expected number of failures when process starts from zero is equal to the expected number of completed repairs when process starts from one, i.e. $M_{0F}(t) = M_{1R}(t)$. Therefore $M_{0F}(t)$ may be computed by formula (3.41). Function M_{0R} corresponds to the delayed renewal process with inter-arrivals $D_1 \sim Gamma(m,\lambda)$, $D_1+U_2 \sim Gamma(k,\lambda)$, $D_2+U_3 \sim Gamma(k,\lambda)\ldots$. Proceeding in the same manner as for the previous renewal functions

$$\begin{aligned} M_{0R}(t) &= \int_0^t \sum_{n=1}^{\infty}\frac{\lambda}{((n-1)k+m-1)!}(\lambda x)^{(n-1)k+m-1}e^{-\lambda x}dx \\ &= \int_0^t \lambda e^{-\lambda x}\sum_{n=1}^{\infty}\frac{(\lambda x)^{nk-(k-m+1)}}{(nk-(k-m+1))!}dx \end{aligned}$$

By Theorem 3.5.1

$$M_{0R}(t) = \int_0^t \lambda e^{-\lambda x} \frac{1}{k} \sum_{r=0}^{k-1} \varepsilon^{r(k-m+1)} e^{\lambda x \varepsilon^r} dx$$
$$= \frac{\lambda t}{k} + \frac{1}{k} \sum_{r=1}^{k-1} \lambda \varepsilon^{r(k-m+1)} \int_0^t e^{-\lambda x(1-\varepsilon^r)} dx$$

Therefore the renewal function M_{0R} can be expressed as

$$M_{0R}(t) = \frac{\lambda t}{k} + \frac{1}{k} \sum_{r=1}^{k-1} \frac{\varepsilon^{r(k-m+1)}}{1-\varepsilon^r} \left(1 - e^{-\lambda t(1-\varepsilon^r)}\right) \tag{3.45}$$

Substituting (3.45) and $M_{0F}(t)$ computed from (3.41) in (3.44) we obtain that the availability function A_0 for Gamma alternating process can be expressed as

$$A_0(t) = \frac{1}{k} \sum_{r=1}^{k-1} \frac{\varepsilon^r}{1-\varepsilon^r} \left(1 - e^{-\lambda t(1-\varepsilon^r)}\right) \left((\varepsilon^r)^{k-m} - 1\right) \tag{3.46}$$

Note that the steady state availability does not depend on the scale parameter λ

$$\lim_{t\to\infty} A_1(t) = \lim_{t\to\infty} A_0(t) = \frac{MTTF}{MTTF + MTTR} = \frac{(k-m)/\lambda}{(k-m)/\lambda + m/\lambda} = 1 - \frac{m}{k}$$

The expressions for both availability functions and the renewal functions are very simple and can be easily implemented on a personal computer. Obviously the simplicity is due to the complex number representation (the result is of course real), however most of the modern programming languages like C++ or Matlab can handle complex numbers very efficiently.

Example 3.5.2. *Let us check the behavior of the availability functions for Gamma alternating process for different set of parameters. Figure 4.4 illustrates functions $A_1(t)$ and $A_0(t)$ obtained from formulas (3.43) and (3.46). Black line on all graphs presents the limiting availability computed from expression (2.54). The manner how the functions approach the limit depends on the shape of the distribution of up and down time. Our purpose is to model situations in which up time is much bigger than down time. For instance the graph on the left upper corner presents the availability functions for the case when uptime is only four times bigger than the down time, what results in relatively low steady state 0.8. In order to increase the steady state availability one has to increase the difference between k and m. However by increasing the shape parameter k the dispersion of the distribution becomes smaller what leads to higher oscillations in the availability functions. Therefore for realistic situations (with high availability) Gamma alternating process may be used only when the oscillatory behavior is expected. On the other hand Gamma distribution is very flexible with regards to the shape and scale what allows to model many realistic cases.*

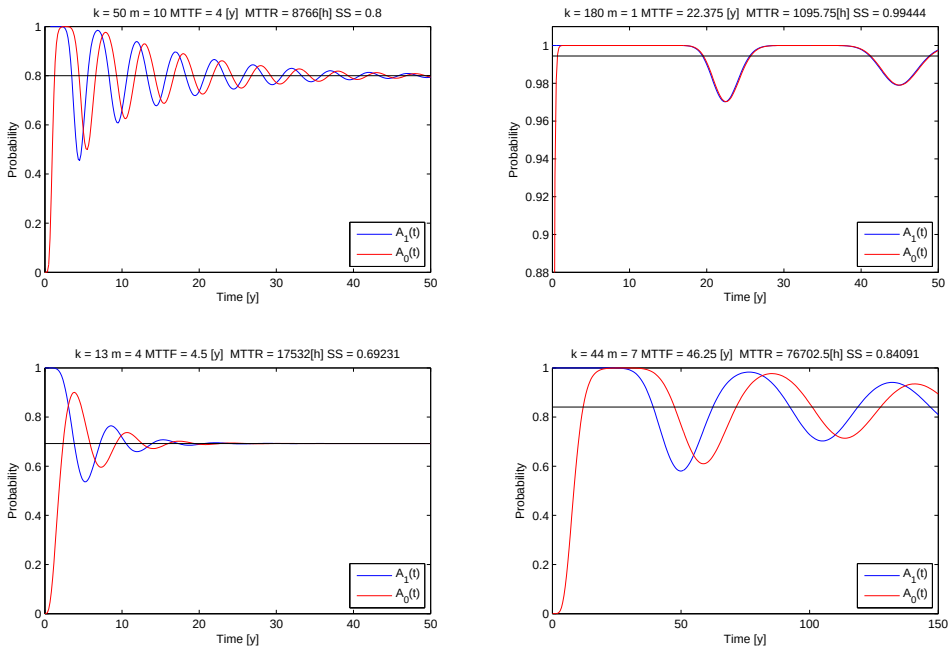


Figure 3.9: Example availability functions for Gamma alternating renewal process.

3.5.2 Testing procedure

It is true that comparison of the methods based on one or two numerical examples can be misleading. An overall picture of the performance of the method depends on its behavior over a wide class of problems. In order to have an objective picture of the performance of the methods we consider a set of problems which have known solutions and which vary in difficulty from easy to hard. Each problem is solved by the routine under consideration, many times with different input parameters and with different number of grid points N. Certain data, such as the time taken and the actual error are collected at the end of each run. Finally, these values are averaged out so as to give an indication of how the routine would cope with a typical problem. These average values are then compared with values obtained using other routines which have undergone exactly the same procedure.

Therefore to test the efficiency of a method with particular problem family we carry out the following procedure with a number of different values of N:

1. run the method with many different values of parameters p_i between the limits $a_i < p_i < b_i$. The parameters p_i are sampled uniformly from the interval (a_i, b_i);

2. after each indvidual run save the error and time taken

3. average out the errors and times. These average values are then plotted against N.

For the each test we used sample of 50 different parameters. To have an objective overview of the performance of the methods it is important to use the same criterion

for evaluating the error. We use the following:

$$\text{Error} = \frac{1}{N}\left(\sum_{i=1}^{N} |r_t(t_i) - \hat{r}(t_i)|^2\right)^{\frac{1}{2}}$$

where $r_t(t_i)$ is the true and $\hat{r}(t_i)$ is the approximate solution at t_i. This is the root mean squre error and may be interperted as a 'distance' between the two vectors.

We give the results for different problem families using five routines which were introduced in Sections 3.3 and 3.4:

- TT – Trapezoid rule for Stieltjes integrals introduced in Section 3.4
- TS – Simpsons rule for Stieltjes integrals also introduced in Section 3.4
- BND – The approximation obtained as the mid point between the bounds introduced in Section 3.3. For the renewal function the bounds are constructed by the recursive formulas (3.10) and (3.11). When the availability function is considered the bounds are obtained by the Method 2 introduced in Section 3.3.2. (Method 1 was not feasible for high steady states).
- MID – Midpoint method introduced in Section 3.4.2
- RK-4 – Fourth order Runge-Kutta method mentioned in Section 3.4.3.

Let us remaind again that these methods approximate function $r(t)$ from the following integral equation

$$r(t) = H(t) + \int_0^t r(t-x)dF_{U+D}(x)$$

(except RK-4 which uses f_{U+D} instead) assuming that H and F_{U+D} are known. However, in most of the situations those input functions have to be approximated as well. Therefore, in checking the performance of the methods we first decide to compare the methods for cases for which H, F_{U+D} are known and later to check the influence on the approximated solution when the input functions contain some error.

3.5.3 Results

Recall that we are interested in the renewal functions that count the expected number of failures in the alternating renewal process: M_{1F}, M_{0F}, M^D_{1F}, M^D_{0F}. Functions M_{1F}, M^D_{1F}, M^D_{0F} correspond to the delayed renewal processes and function M_{0F} is obtained from the ordinary renewal process, for details see Chapter 2. Therefore we decided to perform the tests on the two classes of renewal functions: delayed and not delayed versions (Families 1-4). The performance of the algorithms is also tested for the availability functions A_1 and A_0 (Families 5-9). [3]
Since in practice the most often used distribution for up-time is Weibull we construct the problem families that are close to the cases that may arise when using Weibull distribution by using Gamma renewal process introduced in Section 3.5.1 for which the

[3] For functions A^D_1 and A^D_0 the performance will be similar and it will mainly depend on the accuracy of H (see Table 2.4), if it has to be approximated

analytical solution is known.
Below are the families of test problems for the renewal and availability functions that vary in difficulty from easy to hard:

Family 1: *Renewal function, not delayed, not oscillating*. Renewal distribution $Gamma(k,\lambda)$ with $8 < k < 20$ and $2 < \lambda < 4$.

Family 2: *Renewal function, not delayed, oscillating*. Renewal distribution $Gamma(k,\lambda)$ with $180 < k < 220$ and $6 < \lambda < 12$.

Family 3: *Renewal function, delayed, not oscillating* Renewal distribution $Gamma(k,\lambda)$ with $17 < k < 35$, $m = 1$ and $7 < \lambda < 13$.

Family 4: *Renewal function, delayed, oscillating* Renewal distribution $Gamma(k,\lambda)$ with $30 < k < 70$ and $3 < \lambda < 4$.

Family 5: *Availability function A_1, not oscillating, low steady-state*. The exponential-exponential case with underlying distribution given by (3.55) with the parameters $10 < MTTF < 25$ and $0.39 < MTTR < 0.45$.

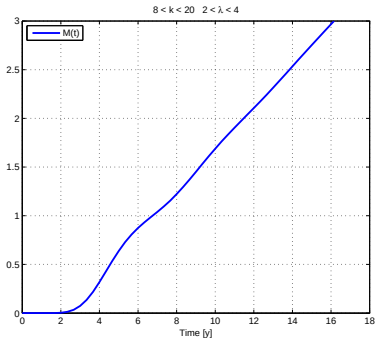
Family 6: *Availability function A_1, low steady-state*. Up-times $U_i \sim Gamma(k-m,\lambda)$, down-times $D_i \sim Gamma(m,\lambda)$, with $20 < k < 30$, $m = 1$ and $3 < \lambda < 7$

Family 7: *Availability function A_1, high steady-state*. Up-times $U_i \sim Gamma(k-m,\lambda)$, down-times $D_i \sim Gamma(m,\lambda)$, with $800 < k < 1050$, $1 < m < 5$ and $10 < \lambda < 15$

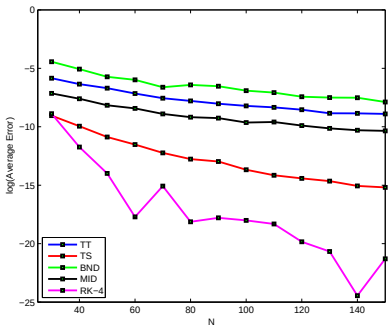
Family 8: *Availability function A_0, low steady-state*. Up-times $U_i \sim Gamma(k-m,\lambda)$, down-times $D_i \sim Gamma(m,\lambda)$, with $50 < k < 80$, $1 < m < 5$ and $10 < \lambda < 15$

Family 9: *Availability function A_0, high steady-state*. Up-times $U_i \sim Gamma(k-m,\lambda)$, down-times $D_i \sim Gamma(m,\lambda)$, with $400 < k < 450$, $1 < m < 3$ and $30 < \lambda < 35$

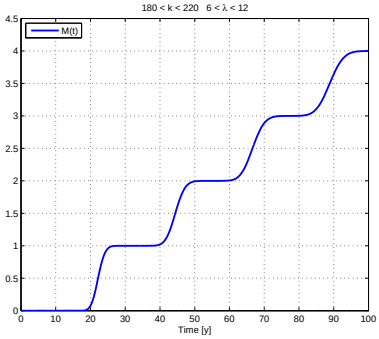
Figures 3.10-3.12 show the achieved average accuracy against N, on a logarithmic scale, for the problem families tested. We now comment on these results.



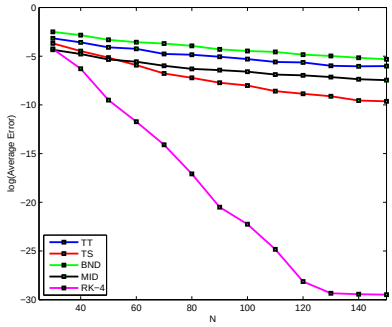
(a) Example renewal function from Family 1.



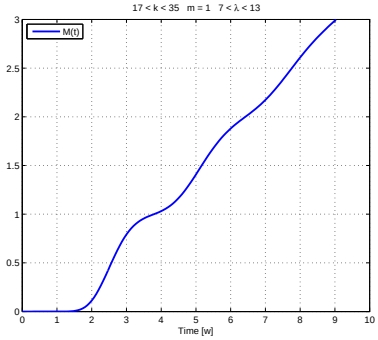
(b) Accuracy for Family 1.



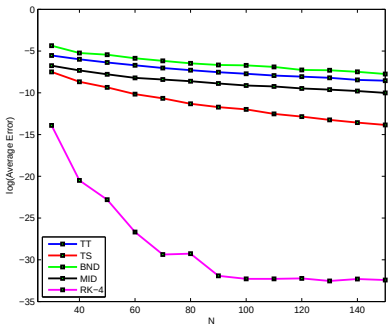
(c) Example renewal function from Family 2.



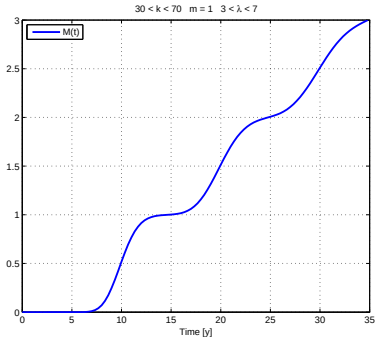
(d) Accuracy for Family 2.



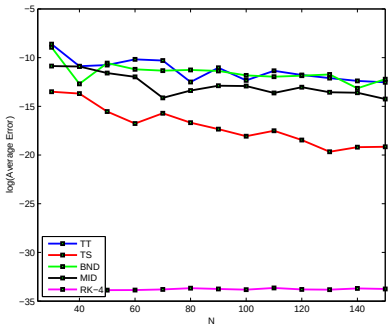
(e) Example renewal function from Family 3.



(f) Accuracy for Family 3.

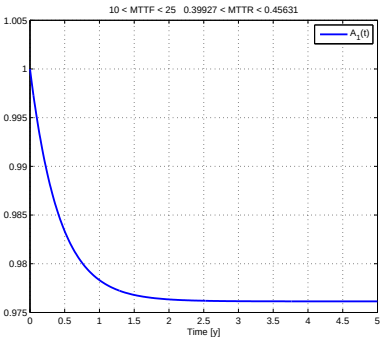


(g) Example renewal function from Family 4.

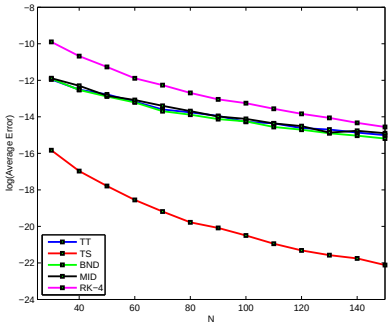


(h) Accuracy for Family 4.

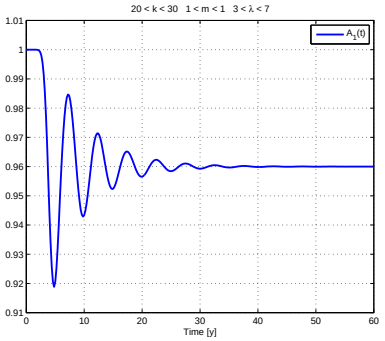
Figure 3.10: Comparison of the methods without computing convolution.



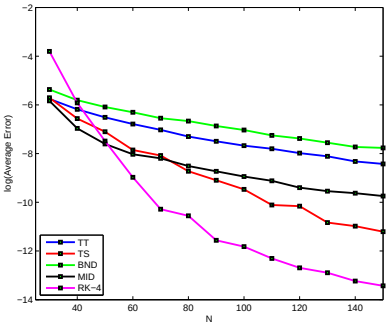
(a) Example availability function from Family 5.



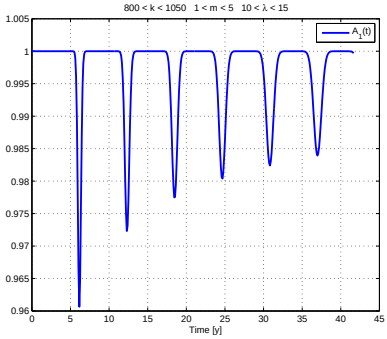
(b) Accuracy for Family 5.



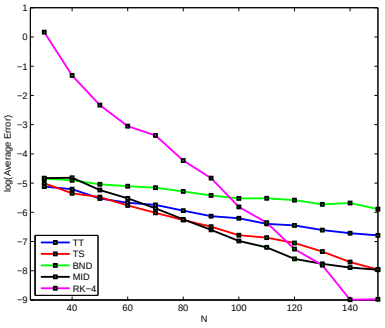
(c) Example availability function from Family 6.



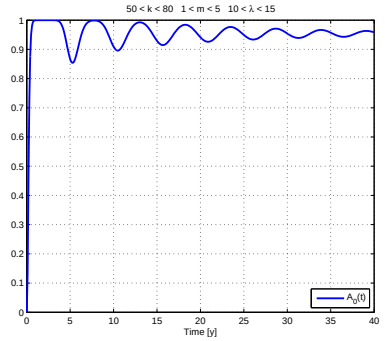
(d) Accuracy for Family 6.



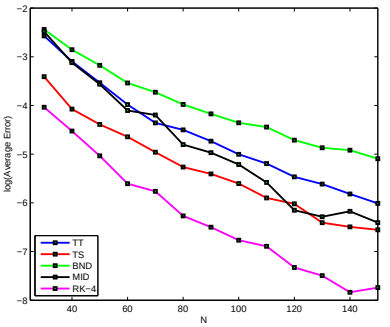
(e) Example availability function from Family 7.



(f) Accuracy for Family 7.

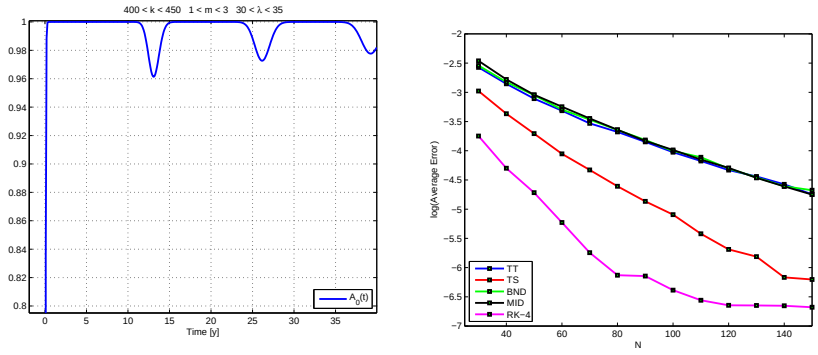


(g) Example availability function from Family 8.



(h) Accuracy for Family 8.

Figure 3.11: Comparison of the methods without computing convolution



(a) Example availability function from Family 9.

(b) Accuracy for Family 9.

Figure 3.12: Comparison of the methods without computing convolution.

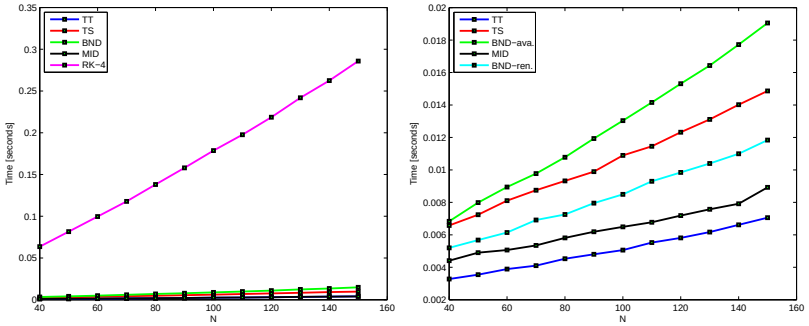
Accuracy

As expected the BND method is the least successful for all families. In general the methods based on the discrete approximation of the continuous quantities suffer from lack of accuracy. This method is also not the fastest since it requires to compute two bounds in case of the renewal function and four bounds in case of the availability function. It appears that TT routine is more accurate on the range of N considered, however the difference is not very significant in case of the renewal functions. This means that it is more efficient to use simple trapezoid rule than the discrete approximation. On the other hand the bad performance of BND is recompensed by returning the error estimate which is missing in other routines. The MID method, that actually works on the grid with $2N$ points, is slightly better than TT for the renewal functions and significantly better for the availability functions, especially for function A_1. It is interesting that all three schemes BND, TT and MID produce almost the same error curves for Family 5, and also similar situation is observed for Family 9. This suggests that for 'easy' availability functions there is no difference in using those three routines and the one with shortest computation time should be chosen (TT method is the fastest since it requires the smallest number of function evaluations, see also Table 3.5.3). Albeit in this case all three methods were outperformed by TS method. It is clearly visible that TS and RK-4 methods are the most accurate for all Families (with exception for RK-4 method for 'easy' Family 5). This is due to the fact that those methods are based on the higher order schemes. The RK-4 method is definitely the most accurate with the fastest speed of convergence, however in some cases it requires more points to be the winner, like for instance for Family 7.

Timings

Results are presented in Figure 3.13. We note that when H, F_{U+D} are known the time taken to approximate r for fixed N is independent of the family and of the problem parameters for all routines except for the BND method which requires to compute 2 bounds for the renewal function and 4 bounds for the availability function. Thus the

method with bounds is twice slower for computing the availability functions than in case of renewal functions. Undoubtedly the RK-4 method is the overall looser with



(a) Elapsed time. (b) Elapsed time without RK-4.

Figure 3.13: Comparison of the methods with respect to the speed.

respect to the speed. The time taken for this method cannot be even compared with other methods on the same scale. The second order TS method is comparable with other methods however it is almost the slowest; although it wins with BND method when the availability functions have to be computed. There is a little difference between two the fastest methods: TT and MID. MID is slightly slower since it requires the values of the distribution on more number of points, however if the distribution is easily accessible this difference is not very significant. In general all methods make the calculations relatively fast (except RK-4 method) on the range of N considered. Nevertheless this small differences in timings may become apparent if the underlying distribution has to be approximated. This is investigated in the next section.

Additional cost related to the computation of convolution

In previous section we have compared the speed and accuracy of different methods for approximating renewal and availability functions. For all considered cases we have used the analytical form for the renewal distribution F_{U+D}. Obviously, if F_{U+D} is unknown then the speed and accuracy of the methods will be affected by additional cost related to the approximation of F_{U+D}. As far as the accuracy is considered it has to be noted that the additional error related to F_{U+D} will not change the rankings of the methods obtained in case when F_{U+D} is error free. This can be explained by the fact that all the methods take as an input the same imperfect distribution. The additional time needed for computing F_{U+D} will depend on the number of grid points on which the value of F_{U+D} is required. The number of F_{U+D} evaluations varies for different methods and different functions. Assume that the solution is required on the grid with N equally spaced points. Table 3.5.3 presents the number of function F_{U+D} evaluations for different methods. The trapezoid rule and the method with bounds are the cheapest in this sense [4]. The midpoint method requires different number of

[4] Method with bounds requires to compute 2 (for renewal function) or 4 bounds (for availability) but all values are computed on the same grid, thus the value of F_{U+D} may be stored and used for computing all bounds

	Renewal fun.	Availability fun.
TT	N	N
BND	N	N
TS	$N+N/2+K$	$N+N/2+K$
MID	$2N$ (N for delayed)	$2N$
RK-4	$3N$	$3N$

Table 3.1: Number of function F_{U+D} evaluations for different methods.

values of the distribution depending on which function is calculated[5]. The Simpson's rule developed by Tortorella needs the values of F_{U+D} on the main grid (N) and also between the every second point (N/2) and also for ascending descending procedure (K) (for details see Figure 3.4.1). The most expensive is the RK-4 method which gives another reason for not accepting it.

3.5.4 Conclusions

We can now try to make an overall assessment of these results. The aim of this comparison was to select the method, perhaps few methods, that are suitable for computing renewal and availability functions. The selected algorithms are going to be implemented in a software that is used for real time calculations and also for large scale models. Therefore the best method should have a good balance between the speed and accuracy.
The winner is TS method since it is the most accurate amongst the fastest methods. Although it works on a finer grid than TT or BND it is still acceptable with regards to the speed. Methods TT, BND are faster but they are not sufficiently accurate what in case of the availability functions is very important. However the MID method is a serious competitor for TS method in case of the availability functions A_1. Albeit the MID point methods looses when the distribution of the process F_{U+D} is not analytical since it needs more points than TS method. The methods with bounds are the least successful with respect to the accuracy. The simple TT method outperforms the BND method in the speed since basically both methods do the same kind of iterations but BND method needs to do it twice in case of the renewal function and four times in case of the availability function. This comparison also shows that there are very accurate methods like RK-4 that come from different theory, however for real applications they are rejected since the time taken for computations is few times longer and in case of large scale models such a performance is simply unacceptable.

3.5.5 Error estimation and usage of bounds

Before the TS routine can be used automatically in the software there are some questions that need to be answered. As it was seen in Figures 3.10-3.12 the size of the error for the approximation differs in all families. Therefore the first question is: what

[5]Based on formula 3.32 for the midpoint approximation if the renewal function is required then $r(0) = 0$ and for delayed renewal process $H \neq F_{U+D}$ thus the value of $F_{U+D}(t_m)$ is not needed. For the availability $r(0) = 1$ and both $F_{U+D}(t_m)$ and $F_{U+D}(t_{m-0.5})$ are required

should be the step-size for the method in order to have desired accuracy? Another problem is related to the computation of convolution, namely: how the error introduced in computing convolution will influence the approximation obtained by the method? We will try to answer this questions in this section.

Choice of the step size

The choice of the step size can be determined in three ways: (1) based on the analytical bound on the error (if such a bound is available), (2) based on the bound on the solution, (3) numerically by investigating the speed of convergence of the method when the number of grid points increases.
In general the studies related to the error analysis are difficult and very often it is not possible to derive a closed formula for the bound on the error of a numerical method. In his paper Tortorella provides analytical bound on the error only in case of the trapezoidal rule and claims (based on test examples) that Simpson's rule is more accurate with the same mesh grid. The bound on error is increasing with time, thus analyzing the error, he also focuses only on last grid point t_N. Based on the bound he suggests the following choice of step size: given a desired maximum error ε for time step t_N the choice of the step size should be chosen of the order $\varepsilon^{1/2}$. We will show that this suggestion is not the most efficient in case of the renewal and availability functions study, mainly due to two reasons: the error is decreasing[6] and the bound on error given in terms of orders of magnitude is too wide.
Let us investigate the actual behavior of the error for the approximation obtained by TS method for renewal and availability functions. Figure 3.14(b) shows that the error becomes smaller (damped oscillations) as the functions approach the steady state and also converges to some limit (in these examples the limit is close to zero). This is expected behavior in case of the renewal type integral equations and it can be explained intuitively as follows. Let $\hat{r}(t)$ denote the approximation to $r(t)$. Assume that $\hat{r}(t)$ satisfies the same integral equation as $r(t)$ but with additional drive term $b(t)$ that introduces the difference between $r(t)$ and $\hat{r}(t)$, then:

$$\hat{r}(t) = H(t) + \int_0^t \hat{r}(t-x)dF(x) + b(t) \tag{3.47}$$

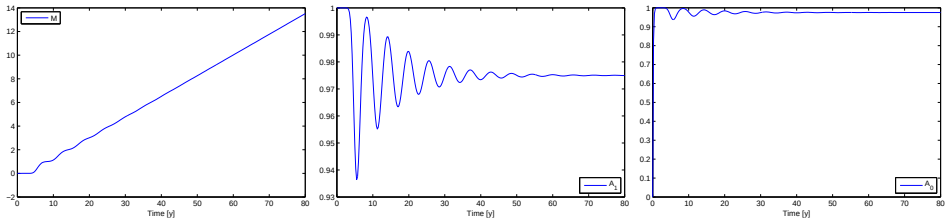
Subtracting this equation from (3.26) gives

$$r(t) - \hat{r}(t) = b(t) + \int_0^t (r(t-x) - \hat{r}(t-x))\, dF(x) \tag{3.48}$$

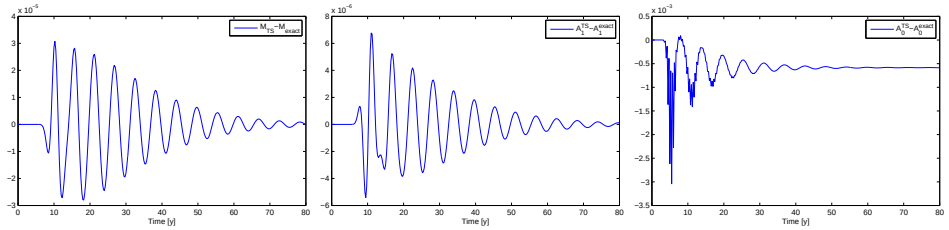
This means that the difference between the approximation and the solution also satisfies renewal type equation. Certainly nothing can be said about the function b, however if one assumes that b is directly Riemann integrable then based on the Theorem 2.3.3 the difference between the approximation and the solution converges to

$$\lim_{t\to\infty} (r(t-x) - \hat{r}(t-x)) = \frac{1}{\mu}\int_0^\infty b(x)dx$$

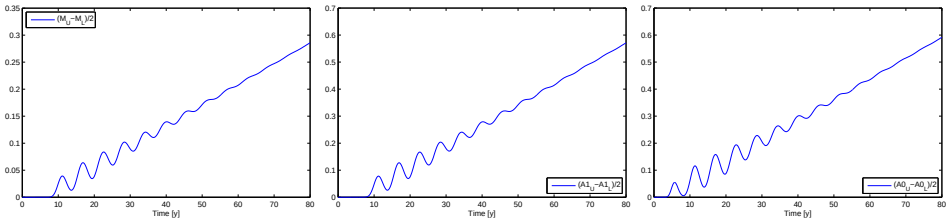
[6] It is non-monotone but it converges to 'zero', see Figure [?]



(a) Example renewal and availability functions.



(b) Actual error between the exact solution and the approximation obtained by TS method.

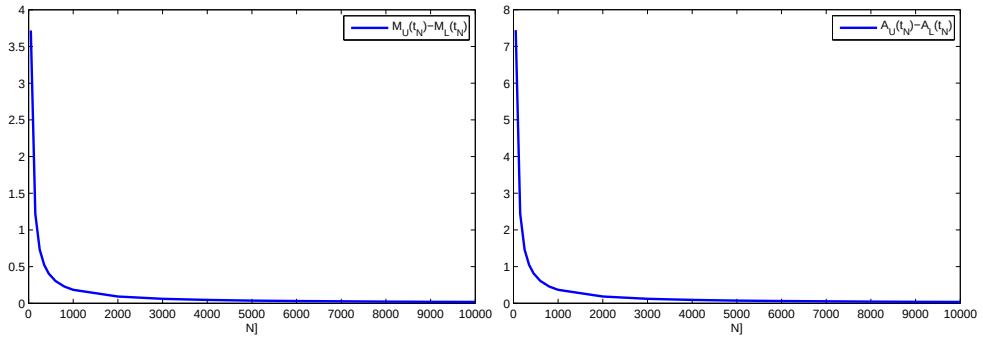


(c) Error estimated by the bounds

Figure 3.14: Example functions for Gamma alternating process with $k = 40$, $m = 1$, $\lambda = 7$; the absolute difference between the approximation obtained by TS and the exact values and error estimation by the BND method. Number of grid points used $N = 321$, steady state $SS = 0.975$.

This reasoning is of course not strict and may be not true in general, however it gives some indication why the error in approximating renewal type integral equations may be not increasing. Therefore in practice the usage of increasing bound on error is not the most efficient method for determining the step size.

In Section 3.3 we have introduced the approximation to $r(t)$ based on the bounds on $r(t)$. In theory this approach is very valuable regarding the error analysis, however the practical usage of this error estimation is very limited and not efficient since the bounds become wider as time increases and also tey require many points to become tight. Therefore the error estimation also increases (example bounds on error are presented in Figure 3.14(c)). Those bounds become tighter as the step size decreases, but it can be shown by examples that the speed of convergence of the bounds to the solution with respect to the step size is slow. Figure 3.15 shows the distance between the upper and lower bound for the renewal and availability functions computed at the last grid point. The convergence is fast but only in a range $\approx [0,1000]$ points, but even in this interval the estimated error (or the distance between the bounds) is relatively big. For instance for availability it is better to use [0,1] as a bound instead of BND method when less than 400 points is used. Therefore we conclude that the usage of bounds as



(a) The renewal function. (b) The availability function

Figure 3.15: Distance between the upper and lower bound for the last grid point.

the error estimate is impractical and in most of the situations it is not possible to use large number of grid points to get such a weak error estimate.

Remark 3.5.3. In some situations it is important to predict the component's availability only in a short time interval, for instance to estimate the time when will be the first possible loss of production due to the next unit's failure. Then the bounds may be used since on the interval of the first renewal ($\approx [0, MTTF + MTTR]$) the bounds are very tight even with small number of grid points.

Remark 3.5.4. Let us compare the approximation given by TS method with the bounds on the availability function A_1. We use the same parameters for the renewal process as before. Figure 3.16 shows that the approximation given by TS method agrees with the bounds. It appears that all 500 points on which A_1 is computed are between the bounds. The calculation of the approximation by TS method took 0.07 second (N =

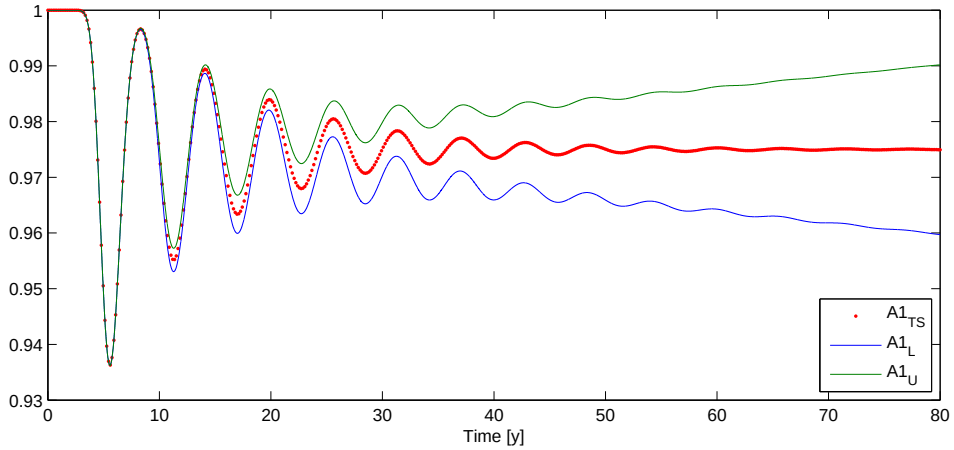


Figure 3.16: Comparison of the bounds on the availability with the approximation obtained by TS method. Number of grid points: for bounds N = 12000 and for TS method N = 500

500) whereas for bounds it was 34.8 seconds (N = 12000).

	$0 < d \leq 3$	$3 < d \leq 5$	$d > 8$
$\max(A_1^{h_0} - A_1^{h_1})$	$10^{-5} - 10^{-7}$	$10^{-7} - 10^{-9}$	$10^{-9} \ldots$

Table 3.2: Suggestion for the initial step size for A_1 with Weibull uptime.

Since it is not possible to find a reasonable error estimate based on the bounds the only way in which results correct to some prescribed accuracy may be obtained is to re-solve the original problem recursively each time using a larger value of N, until two consecutive sets of approximate function values agree to the required accuracy. (Usually the value of N is doubled, that is the steplength is halved, since this enables previously computed values of F_{U+D} to be used again.) The question is what should be the initial step-size so that the number of recursions is small. Certainly the choice of the initial step size will depend on the problem family and parameters. For instance Figure 3.14(b) shows that the accuracy for the function A_0 is much smaller compared to the approximation of the function A_1 on the same grid. We propose the following choice of the initial step size.
Since usually the uptime is much smaller than down time the influence on the shape of the function will be mainly by the distribution of uptime. In our case we are interested in Weibull up-time. The step size should be determined based on the shape of the function; if the function is oscillating then the step-size should be smaller. Let us first focus on the availability function A_1. When the uptime has Weibull distribution the amplitude of the oscillation can be modeled by the shape parameter β, therefore the choice of the step size should depend on this parameter. On the other hand it should also depend on the scale of the problem, thus MTTF should also be included. In the initial choice of the step size one should also have a possibility to control the accuracy by some input parameter. Let d be some positive real number. Based on the experiments with function A_1 we propose to use the following choice of the initial step size

$$h_0 = \frac{MTTF}{d^2 + \beta} \tag{3.49}$$

The variable d is used to specify the desired accuracy for the difference between the approximation obtained on the grid with step h_0 and the approximation obtained on the grid $h_1 = \frac{h_0}{2}$. Table 3.5.5 presents the relation beteween the parameter d and the difference in consecutive approximations that was obtained experimentally. It has to be stressed that this is just the indication of the initial step size and probably based on more simulations one can find more accurate estimate of h_0.
The step size for function A_0 has to be much smaller than the for function A_1 since for realistic parameters A_0 is changing rapidly from zero to almost one on a very short interval, see for example Figure. Therefore we suggest the following choice of the step-size for the initial step for A_0

$$h_0 = \frac{MTTR}{e^2 + \beta} \tag{3.50}$$

This is just a suggestion for further investigation.

Impact of the convolution error on the approximation

So far we have investigated the behavior of the approximation assuming that the the distribution of the sum $U_i + D_i$ is known. As it was mentioned in Section 3.5.3 the time of the computations of r will depend on the number of F_{U+D} evaluations (for TS method it is $N + N/2 + K$). Next to the time it is also important to investigate the influence of the error introduced in approximating F_{U+D} on the overall approximation obtained by TS method. In this section we will try to answer the following question: given the required accuracy on r what should be the accuracy for computing the convolution of $U_i + D_i$? In other words how the error introduced in convolution will change the approximation.
Let $r_\eta(t)$ denote the approximation to the solution of the renewal integral equation obtained by TS method with convolution F_{U+D} computed with accuracy η. Figure 3.17 illustrates the influence of η on the approximation in case of the availability function. It can be observed that the change of η by one order of magnitude changes the approx-

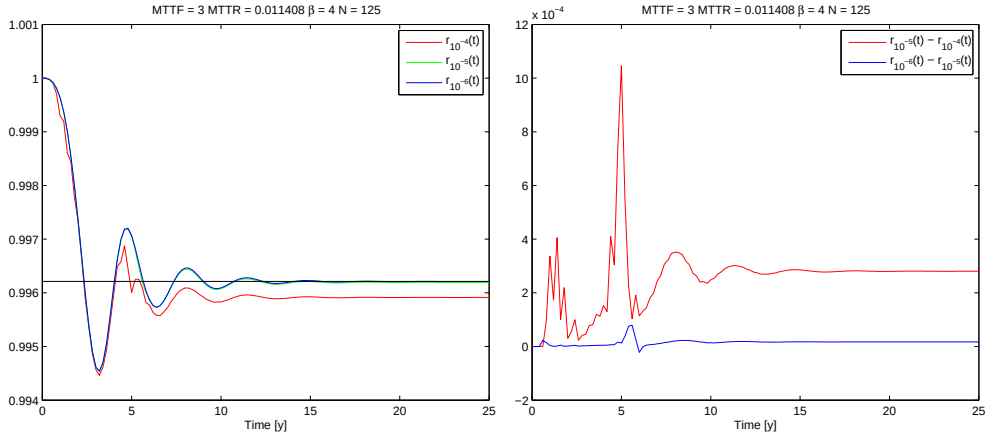


Figure 3.17: The influence of the error in convolution on the availability function A_1, Weibull-Exponential case.

imation also by one order of magnitude. This suggests that the error in the convolution is bigger than the error in the approximation of r. Based on the simulations and comparisons the conclusion is that that the convolution should be approximated with the accuracy at least the same as the desired accuracy for r but not smaller than $10^(-6)$.

3.6 Computing Convolution

All methods for approximating renewal and availability functions require as an input the distribution of the sum of two random variables. In case of A_1 and A_0 it is the convolution of the up and down time, whereas for delayed versions of A_1 and A_0 one also needs to provide the convolution involving a distribution of the time to first event (for details see Section 2.3). Recall that the distribution of the sum of two non-negative independent random variables may be written as (we use notation for lifetimes and

repair times):

$$F_{U+D}(t) = \int_0^t F_U(t-x)dF_D(x) \tag{3.51}$$

$$= \int_0^t F_D(t-x)dF_U(x) \tag{3.52}$$

$$= \int_0^t F_U(t-x)\, f_D(x)dx \tag{3.53}$$

$$= \int_0^t F_D(t-x)\, f_U(x)dx \tag{3.54}$$

Closed form expressions for the convolution are available in only a few special cases. The closure of the normal and gamma families of CDF's under convolution is well known. Other examples are available. From the distributions that are of our interest only two cases are analytically tractable: exponential-exponential and general lifetime - constant down time. If U_i and D_i are both exponentially distributed with parameters λ and μ respectively then

$$F_{U+D}(t) = 1 - e^{-\mu t} + \mu\left(\frac{e^{-\lambda t} - e^{-\mu t}}{\lambda - \mu}\right) \tag{3.55}$$

This is so called *hypo-exponential* distribution. It is also easy to show that convolution of any lifetime distribution with constant leads to shift of the distribution. Indeed, if $P(D_i = \tau) = 1$ then

$$F_{U+D}(t) = P(U+D \le t) = P(U \le t-\tau) = \begin{cases} F_U(t-\tau) & t \ge \tau \\ 0 & t < \tau \end{cases} \tag{3.56}$$

Remaining cases have to be treated by numerical methods. There are many different techniques for computing the convolution of two non-negative random variables. First obvious approach is to compute $F_{U+D}(t)$ by direct numerical integration. Since for each time t it requires to approximate different integral on the interval $[0,t]$ this method may be slow. It is also an open question which quadrature rule should be used and also which form of the function is the easiest to integrate. We will discuss it later.
Another approach is via transforms of the density function. If f_X is a density function of a non-negative random variable X then the general transform $\mathcal{T}$ of f_X can be defined as

$$\mathcal{T}[f_X(t);s] = E[e^{a(s)X}] = \int_0^\infty e^{a(s)t}\, f_X(t)dt \tag{3.57}$$

where $a(s)$ is some given function. Specific choice of the function $a(s)$ will determine the specific transform (if $a(s) = -s$ then $\mathcal{T}$ is a Laplace transform; if $a(s) = -is$ then $\mathcal{T}$ becomes a Fourier transform ; if $a(s) = is$ then $\mathcal{T}$ corresponds to characteristic function). Using independence of up time and down time it is easy to verify that

$$\mathcal{T}[f_{U+D}(t);s] = E[e^{a(s)(U_1+D_1)}] = E[e^{a(s)U_1}] \cdot E[e^{a(s)D_1}] = \mathcal{T}[f_U(t);s] \cdot \mathcal{T}[f_D(t);s]$$

Therefore in order to compute the convolution $F_{U+D}(t)$ one can proceed as follows

1. Compute $\mathcal{T}[f_U(t);s] \cdot \mathcal{T}[f_D(t);s]$

2. Compute $f_{U+D}(t) = \mathcal{T}^{-1}[f_U(t);s] \cdot \mathcal{T}[f_D(t);s]$
3. Compute $F_{U+D}(t) = \int_0^t f_{U+D}(x)dx$

In most situations all three steps have to be approximated numerically. In case of Weibull distribution the Laplace transform and characteristic function can be expressed in terms of infinite alternating series. Those representations are presented in Appendix A. However, from the practical point of view the usage of such representations is limited. Mainly there are two reasons for that: 1) the convergence is slow and therefore the number of summed elements have to be large and 2) it is not possible to determine a truncation parameter for the summation because the series are alternating.

Despite the fact that this method involve many numerical issues related to steps 1 and 2 the main reason why we drop it is step 3. We look for a method that approximates $F_{U+D}(t)$ but without using $f_{U+D}(t)$ to avoid additional computational effort, see discussion in Section 3.4.3.

Another approach for numerical convolution of life distributions was proposed by Tortorella in [46]. The author proposes simple recursive formulas for computing $F_{U+D}(t)$ based on trapezoid and Simpson's rule for Stieltjes integrals. The method produce reasonable results for realistic values of parameters and is worth considering.

Another popular method is approximation through discretization of the continuous distributions. The popularity is attributed to the ease of calculations at the cost of reduced accuracy. We will present this approach in more details. There are also other methods for computing convolution but most of them focus on probability density functions that are not of our interest (see for example [22, 2]).

3.6.1 Direct integration

This is very direct approach however it seems to be reasonable in many cases. The procedure is based on approximating Riemann integrals of the form (4.7) and (4.10). Let us first decide which form is easier to integrate based on Weibull - exponential case (for parametrization of the distributions see Appendix A). From (4.7) we have

$$F_{U+D}(t) = \int_0^t \left(1 - e^{-((t-x)/\alpha)^\beta}\right) \mu e^{-\mu x} dx \tag{3.58}$$

$$= F_D(t) - \mu \int_0^t e^{-((t-x)/\alpha)^\beta - \mu x} dx \tag{3.59}$$

and from (4.10)

$$F_{U+D}(t) = \int_0^t \left(1 - e^{-\mu(t-x)}\right) \frac{\beta}{\alpha} (x/\alpha)^{\beta-1} e^{-(x/\alpha)^\beta} dx \tag{3.60}$$

$$= F_U(t) - \beta\alpha^{-\beta} e^{-\mu t} \int_0^t x^{\beta-1} e^{-(x/\alpha)^\beta + \mu x} dx \tag{3.61}$$

Thus actually there are four possibilities. Note that integrant in (3.61) does not dependent on t so for each t we can integrate only increment and add to previous value. Namely

$$\int_0^{t+dt} x^{\beta-1} e^{-(x/\alpha)^\beta + \mu x} dx = \int_0^t x^{\beta-1} e^{-(x/\alpha)^\beta + \mu x} dx + \int_t^{t+dt} x^{\beta-1} e^{-(x/\alpha)^\beta + \mu x} dx$$

This is valuable only in theory. Usually life time is much longer than repair time (see also Section 3.1). For instance we can say that a component which have expected time to failure $MTTF = 3$ years needs to be repaired in average for 10 hours (≈ 0.0011 of the year). With $\beta = 4$ (also realistic value), this corresponds to $\mu = 876.6$ and $\alpha = 3.31$. If we take out constants from the integral as in (3.61) the $e^{-\mu t}$ then we have to calculate big powers of e and after integration multiply by very small number. This leads to numerical problems thus we keep constants under the integral. Eventually we have to decide which form of the integrant in equations (3.58), (3.59), (3.60) is the most convenient for integration. This decision is based on the experiments with different parameters. For plotting purposes let us label the integrants

$$f_{3.58}(x) = \left(1 - e^{-((t-x)/\alpha)^{\beta}}\right) \mu e^{-\mu x}$$

$$f_{3.59}(x) = e^{-((t-x)/\alpha)^{\beta} - \mu x}$$

$$f_{3.60}(x) = \left(1 - e^{-\mu(t-x)}\right) \frac{\beta}{\alpha} (x/\alpha)^{\beta-1} e^{-(x/\alpha)^{\beta}}$$

where α, β, μ and t are fixed parameters. We also want to investigate the behavior of these functions for different t when α, β, μ are fixed. Figure 3.18 illustrates functions $f_{3.58}$, $f_{3.59}$ and $f_{3.60}$ for two sets of parameters and different t. From this Figure it is

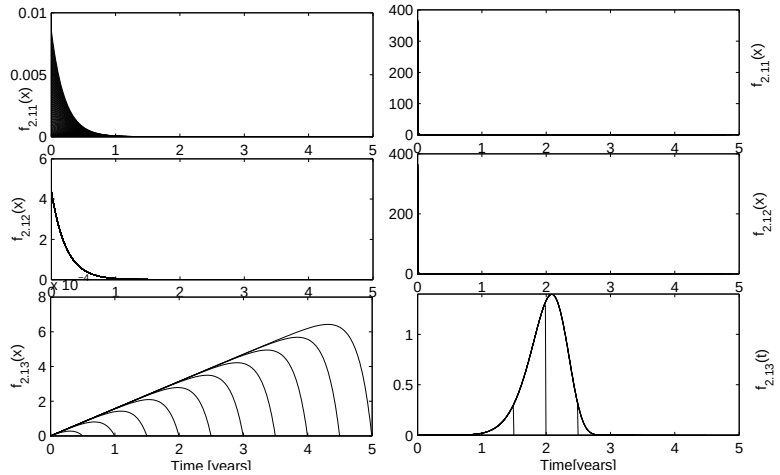


Figure 3.18: Functions for integration in Weibull-exponential case, each line denotes function for different time t. Parameters for distribution: Graphs on the left: $MTTF = 100$ [y], $MTTR = 2000$[h] and $\beta = 2$ what corresponds to $\alpha = 112.83$ and $\mu = 4.38$. Graphs on the right: $MTTF = 2$ [y], $MTTR = 24$[h] and $\beta = 8$ what corresponds to $\alpha = 2.12$ and $\mu = 365.25$.

easy to notice that functions $f_{3.58}$, $f_{3.59}$ are difficult to integrate numerically whereas the function $f_{3.60}$ seems to be quite reasonable. It appears that this is not an isolated case. We checked also other parameters and other distributions and the conclusion is that for realistic values of parameters where life time is much bigger than lifetime the best form of the convolution for numerical integration is

$$F_{U+D}(t) = \int_0^t F_D(t-x) f_U(x) dx = \int_0^t F_D(x) f_U(t-x) dx$$

This information is also important in discrete approach presented in the next section. Now we need to choose a numerical method for approximating the integral. As it can be observed on Figure (3.18) smart selection of the points to the grid for integration may lead to higher speed and accuracy. Simple trapezoidal rule with equidistant grid is unacceptable. We propose to use adaptive Simpson's rule. Details of this method are precisely described in [21] and also Appendix C.

3.6.2 Discrete approximation and bounds

Recall that the aim is to approximate $F_{U+D}(t) = P(U+D \leq t)$ for many values of t. It is possible to approximate continuous convolution by its discrete counterpart. This can be done for instance in the following way. Let us say that we require value of $F_{U+D}(t)$. Choose a grid $0 = t_0 <, \ldots, < t_n = t$. Then

$$F_{U+D}(t_n) = \int_0^{t_n} F_T(t_n - x) dF_U(x)$$

The above integral can be approximated by

$$F_{U+D}(t_n) \approx \sum_{k=1}^{n} F_T(t_n - t_k)\left(F_U(t_k) - F_U(t_{k-1})\right) = \sum_{k=1}^{n} F_T(t_n - t_k) p_k \tag{3.62}$$

where $p_k = F_U(t_k) - F_U(t_{k-1})$. Equation (3.62) is the discrete approximation of the continuous convolution. Note that there is no need to discretize two distributions. To increase the speed of computing this summation one can use so called Fast Fourier Transform (FFT). Evaluating these sum directly would take $O(n^2)$ arithmetical operations. The FFT algorithm computes the same result in only $O(n \log n)$ operations. We stress that the only reason for using FFT method is increasing the speed. Therefore one can compute the discrete convolution $F_{U+D}(t_n)$ very quickly even for large n. For instance for $n = 10000$ it takes only 0.5 second to compute $F_{U+D}(t_n)$ using FFT, wheras direct summation takes about 100 times longer. Such long sequences may be required if the discretization error needs to be decreased. However as it will be seen in Section 3.6.4 the discretization methods have in general limited accuracy.

3.6.3 Newton-Cotes rule for Stieltjes Integrals

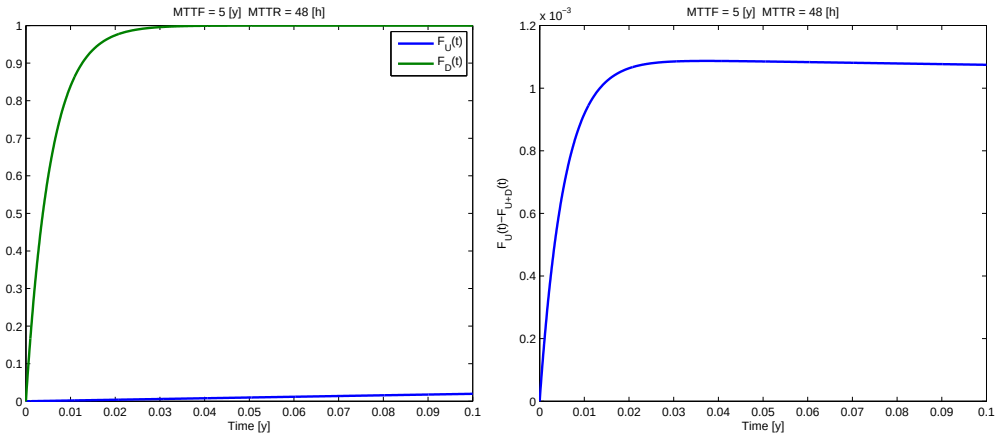
In his paper [46] Tortorella develops two Newton-Cotes rules for Stieltjes Integrals based on trapezoid and Simpson's rule. We will also use this methods for comparisons.

3.6.4 Comparison

Let us give an example distribution functions that needs to be convoluted. Assume that $U_i \sim Exp(\lambda)$ and $D_i \sim Exp(\mu)$. The realistic parameters that may occur in practice are: $MTTF = 5 years$ and $MTTR = 48 hours$. This corresponds to $\lambda = 0.2$ and $\mu = 182.625$. Figure 3.19(a) presents the distributions for this case. It can be observed that the distributions have very different shapes; the distribution of downtime is almost one when the distribution of uptime is almost zero. This is the main reason why most of the numerical methods fail to compute the convolution for the realistic parameters with reasonable accuracy. For realistic parameters the convolution F_{U+D} will be very similar

to the distribution of F_U since the downtime is relatively small. This difference for the example considered is of order of 10^{-3} for small t and decreases to zero as t increases, see Figure 3.19(b). Usually it is required to compute the availability and renewal functions on the interval $[0,t]$ for some large t. Let us assume that the availability function is required on the following grid: $0 < h < \cdots < mh = 20$. Then the values of the distribution F_{U+D} are required also on the same grid. Figure 3.6.4 presents the absolute error for approximating F_{U+D} by the recursive trapeze and Simpson's methods and also by the discrete approximation introduced in Section 3.6.2.

We have used 1000 points for the recursive formulas and 10000 for discrete approximation and this results in step size $h = 0.02$ and $h = 0.002$ respectively. Of course if the step size is bigger than the region where the the distribution F_D is changing then the approximation will result in the error similar to the difference between the F_U and F_{U+D}. From this figure it can also be observed that in case of the discrete approximation even the large number of points will not improve the accuracy. Therefore it is



(a) Distribution of uptime together with distribution of dowtime

(b) Difference between $F_U(t)$ and $F_{U+D}(t)$

Figure 3.19: Example distributions.

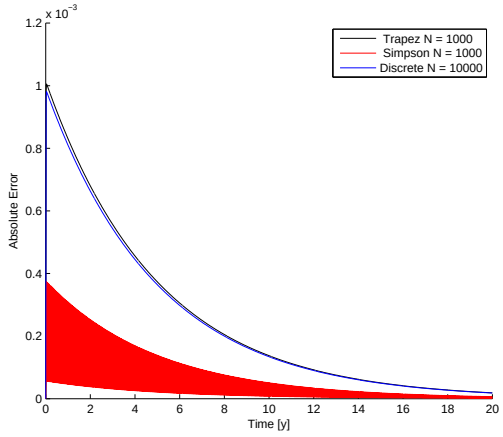


Figure 3.20: Absolute error for computing convolution.

clear that the uniform grid for computing convolution for realistic values of parameters is not the best choice. Another problem arises when the grid $0 < h < \cdots < mh = 20$ and the grid on which F_{U+D} was computed do not match. Then the interpolation is necessary. It is also not straightforward how to use FFT on the non-uniform grid. Ref [23], however, presented some approach for computing the convolution using FFT on non-uniform grids.
It appears that the most suitable approach for computing the convolution is by using direct integration introduced in Section 3.6.1. The advantage of this method is that it approximates the values of $F_{U+D}(t)$ for every t independently and also since it is using adaptive grid it may recognize problematic cases automatically. This method can easily obtain accuracy of order 10^{-7} and higher. The error estimation provided by the method is also reasonable.

3.7 Steady state detection

As it was mentioned in Section 2.3.7 both renewal and availability functions converge to known asymptotic values. From the computational point of view it is very important to investigate the speed of convergence especially when the solution is oscillatory. By steady state we mean the first time t_s after which the function will remain in a distance less than some given ε from its asymptotic value. The methods that estimate the steady state can be divided into two groups: 1) compute t_s before calculating the function and 2) compute the t_s by using the values of the function. Naturally the first methods are more valuable since t_s may be used for choosing a step size in the algorithm for computing the function.

Let us consider the renewal function $M_{1R}(t)$ that counts expected number of completed repairs. According to (2.23) asymptotic expansion in this case is

$$J(t) := \lim_{t\to\infty} M_{1R}(t) = \frac{t}{E[U]+E[D]} - \frac{1}{2} + \frac{Var[U]+Var[D]}{2([U]+E[D])^2}, \;\; t \to \infty$$

Similarly for corresponding the availability function A_1

$$A = \lim_{t\to\infty} A_1(t) = \frac{E[U]}{E[U]+E[D]}$$

From the numerical point of view it is important to know how rapidly the asymptotic result is approached. For each function let us define a relative error that measures the relative distance between the function and asymptotic value

$$E_M(t) = \left|\frac{M_{1R}(t)-J(t)}{M_{1R}(t)}\right| \text{ and } E_A(t) = \left|\frac{A_1(t)-A}{A_1(t)}\right| \tag{3.63}$$

In general it is not possible to find t_s analytically. However it is known that the rapidity with which the oscillations will die out depends on the dispersion of the renewal distribution, see [15]. Therefore we want to find some relation between the dispersion of the distribution and steady state. The simplest measure of the dispersion of a distribution is a *coefficient of variation* defined as

$$CV = \frac{\sigma}{\mu}$$

where σ is standard deviation and μ is the first moment of a distribution.

3.7.1 "Switch-over" point methods

As proposed by [22] or [20] the relation between the CV and the steady state may be found experimentally. Both authors suggest the following similar idea for finding steady state for a renewal function. They propose to use two piece approximations for the renewal function. First piece is approximated by using some numerical method (ref [22] uses Pade approximates and ref [20] uses method of Xie [48]) and second piece is obtained by using linear asymptotic expansion. The point at which the method switch from numerical approximation to linear is called a switch-over point and corresponds to t_s in our interpretation. The procedure for the relation between the CV and t_s (or switch over point) is as follows: fix some desired tolerance level ε, specify parameters to the distribution and compute CV, next compute renewal function for consecutive t's until the relative difference between the asymptotic value and computed renewal function is less than ε. Repeating this procedure for different CV will give a table of points (CV, t_s). The authors disagree how to choose t_s for CV that is not in the table. Ref [22] proposes to use third-order polynomial to interpolate unknown values, however ref [20] gives a counter example for which the method of [22] is incorrect and by himself proposes linear interpolation.
The disadvantage of such approaches is that for each distribution and for each ε one has to have different table. Another difficulty appears with scale parameters of the distribution. Let us take for instance $Weibull(\alpha,\beta)$ to be the renewal distribution. It is easy to show that in this case CV is independent of α. Therefore after estimating the steady state t_s based on the table one has to scale it by using parameter α (this not mentioned in the example of [20]). Scaling is not that obvious in case when the renewal distribution is a convolution of two different distributions.
Recall that in our situation the renewal distribution is the convolution of up time and down time distribution F_U and F_D respectively. If we choose for instance $U_i \sim Weibull(\alpha,\beta)$ and $D_i \sim Exp(\eta)$ then the CV will depend on three parameters $CV = CV(\alpha,\beta,\mu)$. It easy to find two sets of parameters $(\alpha_1,\beta_1,\eta_1)$ and $(\alpha_2,\beta_2,\eta_2)$ for which

$$CV(\alpha_1,\beta_1,\eta_1) \approx CV(\alpha_2,\beta_2,\eta_2)$$

and for which the steady states are different. To exemplify let us take $(\alpha_1,\beta_1,\eta_1) = (5.4,5,168/8766)$ and $(\alpha_2,\beta_2,\eta_2) = (20.4,5,336/8766)$ (parameters α and μ have been scaled by factor of 4 and 2) then in both cases $CV \approx 0.22$. But the steady states are respectively $t_s \approx 96$ and $t_s \approx 383$ (computed with $\varepsilon = 10^{-8}$). For calculating the renewal function in this example we have used TS method. It is not straightforward, as it is in case of Weibull distribution, how to scale the steady states for the cases with the same coefficient of variation. Therefore we do not employ those techniques neither for renewal function nor for the availability function.

3.7.2 Rough estimation

Another much simpler approach for estimating the steady state was proposed by Cox in [15]. Let as before σ and μ denote standard deviation and the first moment of a

renewal distribution. Then he claims that the following estimation of the steady state is applicable when $CV \ll 1$

$$t_s = \frac{\mu^3}{\sigma^2} = \frac{\mu}{CV^2} \tag{3.64}$$

Note that to the contrary with previous approaches the steady state depends not only on the CV but also on μ (it also includes the information about the scale). Although (3.64) is a very rough estimate and it does not provide any information about the error we decide to test it due its simplicity. Let us consider the following example

Example 3.7.1. *Assume that* $U_i \sim Weibull(\alpha, \beta)$ *and* $D_i \sim Exp(\eta)$ *with* $MTTF = 5$ *[years] and* $MTTR = 3$ *[days] and* $\beta = 5$. *This corresponds to* $\alpha = 5.44$ *and* $\eta = 121.75$ *with* $CV = 0.23$. *Steady state estimated in this case by formula (4.1) is* $t_s = 95.76$ *[years]. Figure 3.21 presents both renewal and availability function computed on the interval* $[0,\ 95.76]$. *For computation we used method of [47] with absolute accuracy for computing convolution* $\varepsilon_{conv} = 10^{-8}$ *and with 1000 grid points. The*

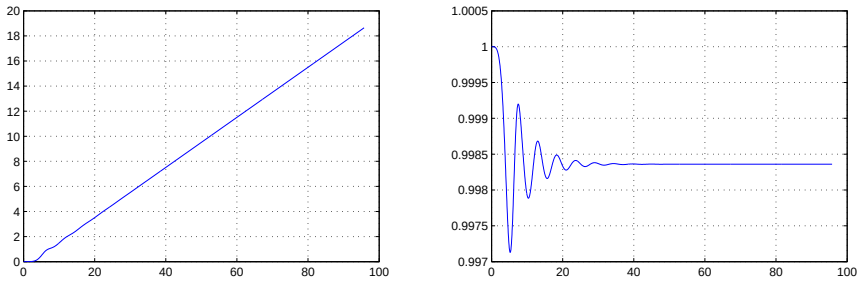


Figure 3.21: Renewal function $M_{1R}(t)$ and availability function $A_1(t)$ from Example 3.7.1

relative difference between the asymptotic value and the approximation for the last grid point is

$$E_M(t_s) = 6.48 \cdot 10^{-8} \text{ and } E_A(t_s) = 5.2 \cdot 10^{-8}$$

If we increase in the example above the accuracy of computing the functions (by either increasing the accuracy of convolution or by increasing the number of grid points) then the values $E_M(t_s)$ and $E_A(t_s)$ become smaller. Similar behavior was observed for other parameters. This means that the steady state estimated by formula (4.1) may be too big as compared to desired accuracy. By visual inspection of Figure 3.21 it may also be observed that the steady state has been reached much earlier than the estimation given by (4.1). In general increasing β (decreasing discrepancy of the distribution) increases the value of t_s. Figure 3.22 illustrates the relationship between β and t_s for Weibull-exponential case with parameters $MTTF = 1$ [y], $MTTR = 1$[h]. Also it is an open question how small should be CV. The conclusion that may be drawn from this approach is that estimation given by Cox is very simple to use but from the practical point of view very often it overestimates the steady-state. The suggestion is that it may be used as an upper bound for the steady state estimation.
Since above methods are not successful in detecting the steady-state one needs to estimate it numerically.

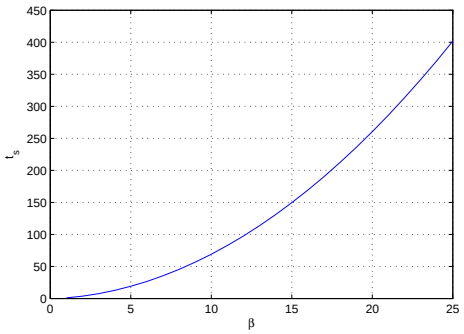


Figure 3.22: Relationship between steady state t_s and β for Weibull-exponential case with parameters $MTTF = 1$ [y], $MTTR = 1$[h]

3.7.3 Numerical

In case of availability function the speed of convergence may be investigated by looking at extremes. The location of the extremes of the availability functions is not a trivial task and in practice those extremes have to be detected numerically during the computations of the function.

3.8 Example availability functions

In this section we give a few examples of the availability functions computed by TS method. For detailed explanation of the meaning of the function see Section 2.3.

Example 3.8.1. Functions A_1. *This function gives the probability that a component is functioning at time t in the future given that today, at time $t = 0$, it has been put into operation. This function is the most often used in the availability studies, especially in the designing phase where the components are assumed to be new. Figures 3.23 and 3.24 illustrate example functions for the case when the uptime has Weibull distribution and downtime is either exponential or lognormal. Parameters for each case are given above the plots: MTTF (mean time to failure in years), MTTR (mean time to repair in hours), β shape parameter for Weibull distribution, SS – steady state availability, N – number of grid points used, step – corresponding step-size in days, time – elapsed time for computing the function including computing convolution, ε_{conv} – accuracy used for computing convolution.*

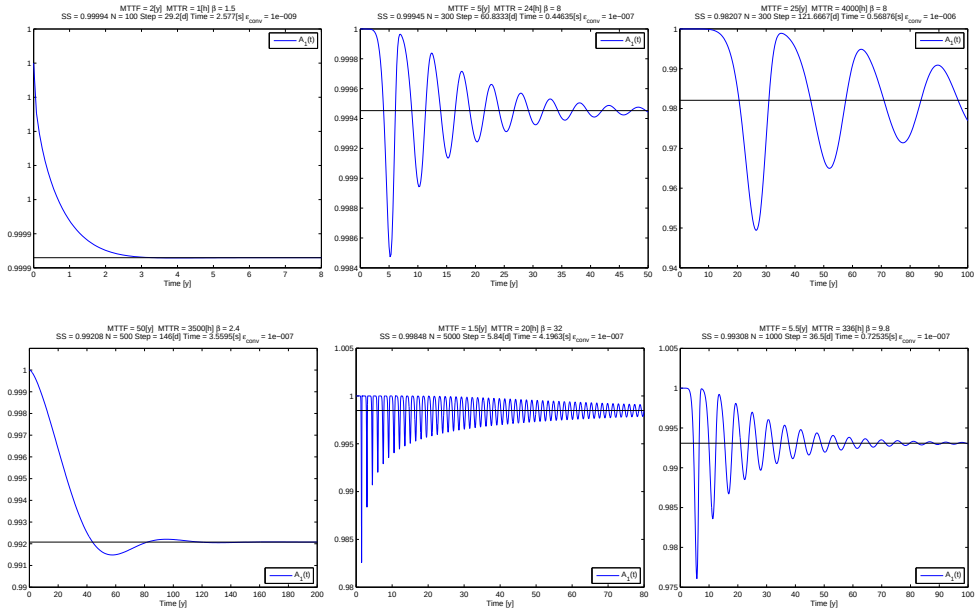


Figure 3.23: Example availability functions A_1 for Weibull-Exponential case.

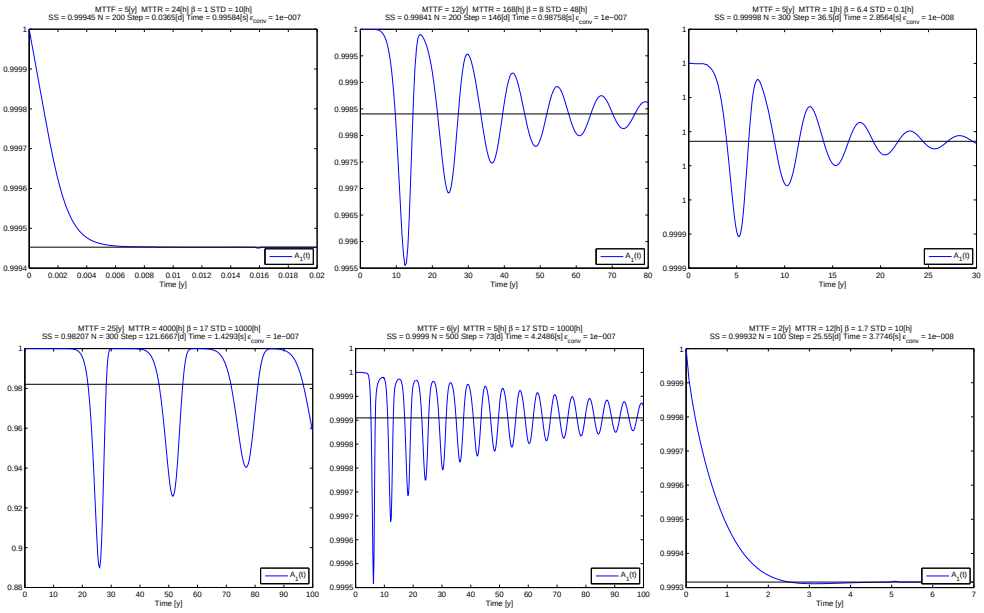


Figure 3.24: Example availability functions A_1 for Weibull-Lognormal case.

Example 3.8.2. Functions A_0. *This function gives the probability that a component is functioning at time t in the future given that today, at time $t = 0$, the repair has just started. Since usually the repair times are much smaller compared to life times function A_0 will be rapidly increasing for small t. From Figure 3.25 it can be observed*

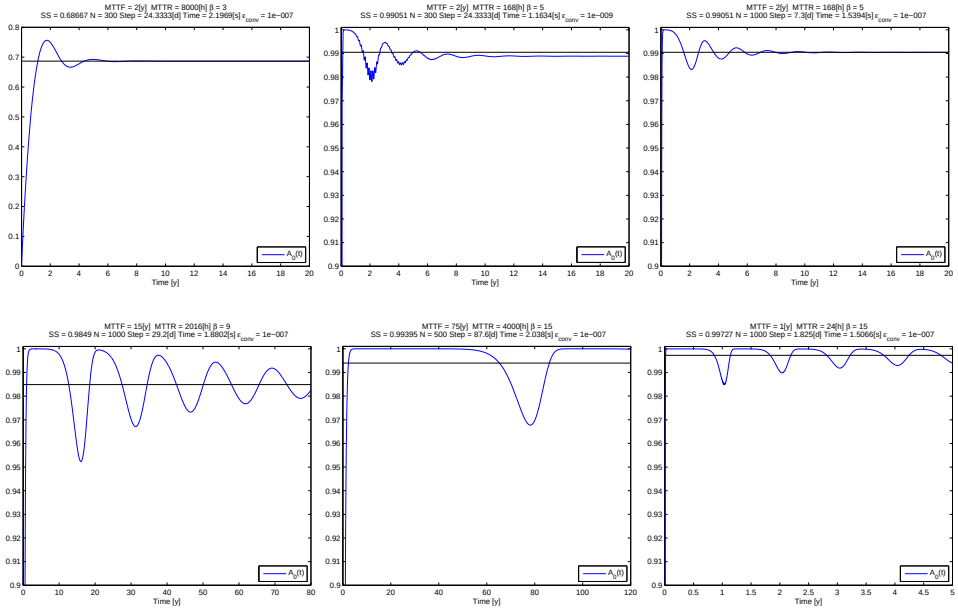


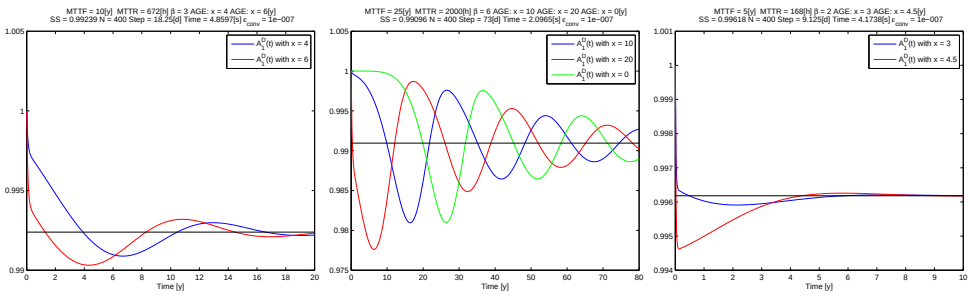
Figure 3.25: Example availability functions A_0 for Weibull-Exponential case.

that much finer grid is required in case A_0 than A_1. Sometimes when the step size is too big the function may not converge to steady state.

Example 3.8.3. Functions A_1^D. *Remind that A_1^D is a delayed version of A_1 and corresponds to the situation in which a component is already functioning for x units of time at $t = 0$. In practice this function is used more often in the operation stage rather than in the design since it incorporates the age of a component. Figure 3.26 illustrates few examples of functions A_1^D. As it can be observed the function A_1^D may be significantly different from A_1. Each graph presents two functions with the same set of parameters but with different age of the component. For instance plot 3.26(a) illustrates the availability of the two components that are 4 and 8 -year old. In general, the computation of A_1^D is more expensive than A_1 since it requires additionally the calculation of the convolution of the remaining lifetime distribution with the repair distribution. The times shown on the Figures are for one function.*

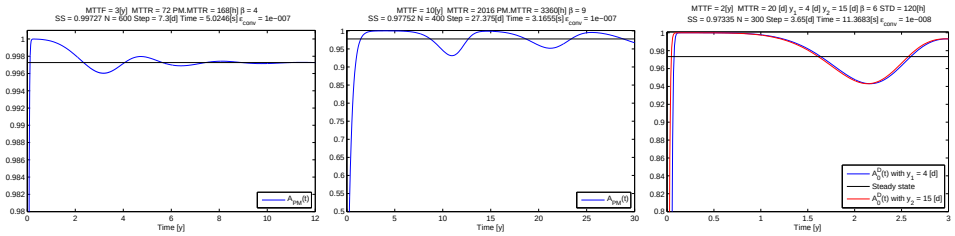
Example 3.8.4. Functions A_0^D. *As mentioned in Section 2.3.6 function A_0^D has two applications: (1) it models the availability of a component that is in repair for y time units at $t = 0$ and (2) it can be used to model a planned maintenance with variable duration that has different distribution than usual repair times. Figure 3.27 illustrates two functions A_{PM} for which the maintenance time is assumed to be exponential with mean PM.MTTR.*

In case when the down time is exponentially distributed the remaining repair has again exponential distribution with the same parameter (memoryless property), there-



(a) (b) (c)

Figure 3.26: Example availability functions A_1^D for Weibull-Exponential case.



(a) A_{PM} for Weibull-Exponential case with exponential maintenance.

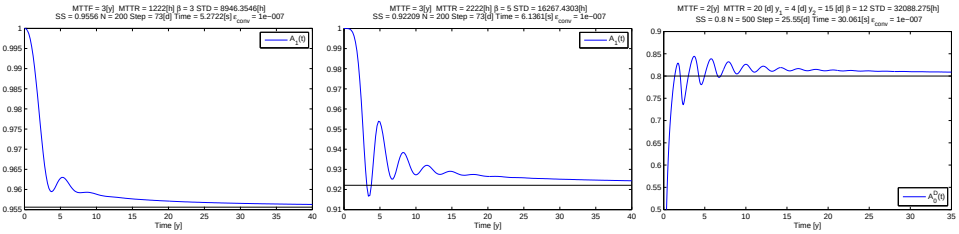
(b) A_{PM} for Weibull-Exponential case with exponential maintenance.

(c) A_0^D for Weibull-Lognormal case with two delays.

Figure 3.27: Example availability functions A_{PM} and A_0^D.

fore $A_0(t) = A_0^D(t)$ for all t. Figure 3.27(c) presents the function A_0^D in case when the repair time has lognormal distribution with MTTR=20 days for two different times of the repair in progress: 4 days (blue) and for 15 days (red).

Example 3.8.5. *Figure 3.28 illustrates some availability functions for not realistic parameters.*



(a) A_1^D for Weibull-Lognormal.

(b) A_1^D for Weibull-Lognormal.

(c) A_0^D for Weibull-Lognormal case.

Figure 3.28: Example availability functions for not realistic parameters.

3.9 Influence of the repair distribution on the availability function

Since in practice the repair time is much smaller as compared to the lifetime of a component it is important to check how the distribution of the repair time influences the availability function. In this section we investigate the difference between the availability functions A_1 for the case when the lifetime has Weibull distribution and when the repair times are either exponential or constant. From the computational point of view it is important to make this sort of comparison since in case when the down time is constant there is no need to approximate the distribution of $U_i + D_i$ (see Section 3.6 for details).
It is clear that when the steady state (SS) availability is high then the repair time is relatively small compared to the lifetime (see (2.54)). And if the repair time is small then the influence on the (availability) function will also be small. Therefore we compare the difference between the availability functions for exponential and constant repair times for three different steady states: 0.95, 0.97 and 0.999. It also appears that this difference depends on the shape of the function itself. In this case, the shape of the availability function may be modeled by β (shape parameter for the Weibull distribution). We decide to compare the difference for $\beta = 2, 8, 15$. Results are presented on Figures 3.29, 3.30 and 3.31. All functions were computed on the grid with $N = 300$ points and for computing the convolution of up and down time in exponential case the accuracy of 10^{-9} was used. Upper plots in the those figures illustrate the availability functions with the up time being Weibull and with the downtime being constant and exponential. The length of the down time for the constant case equals to the MTTR for of the exponential downtime. This match will assure that both functions will converge to the same steady state. Lower plots present the difference between corresponding availability functions. In general, behavior of the availability functions is similar and

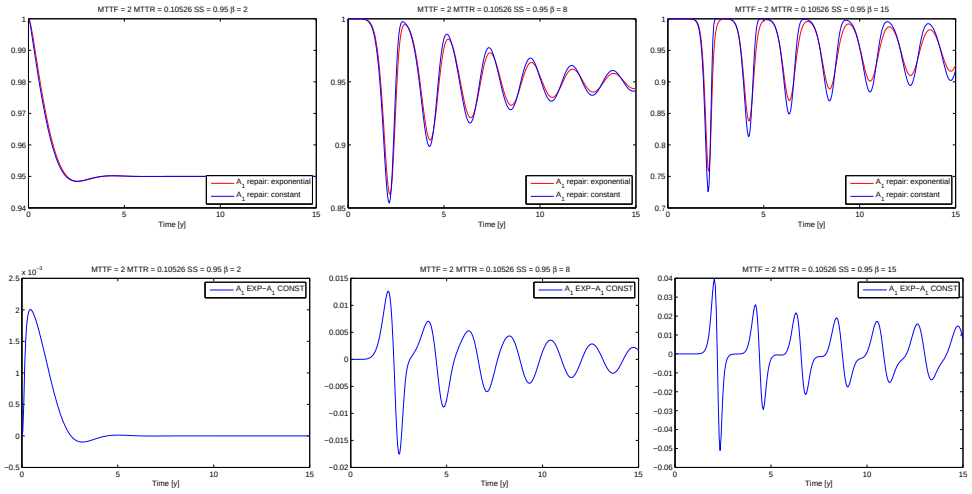


Figure 3.29: Comparison of the availability function for different repair times: steady state 0.95

the difference becomes more significant when the function is oscillating. On Figure

3.29 we observe that the error in using constant repair time instead of exponential one is of order 10^{-3} for non oscillatory behavior and increases to 10^{-2} for case when $\beta = 15$. This error decreases to zero as the functions converge to the steady state. When the steady state is higher the error becomes smaller. For instance for $SS = 0.999$

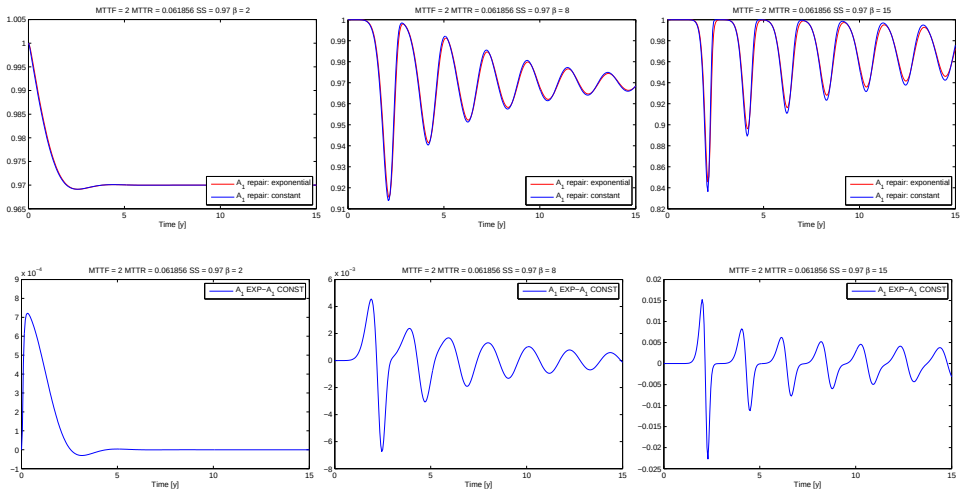


Figure 3.30: Comparison of the availability function for different repair times: steady state 0.97

the difference between the availability functions ranges from 10^{-7} to 10^{-5} depending on their shape. The conclusions that may be drawn from this investigations are:

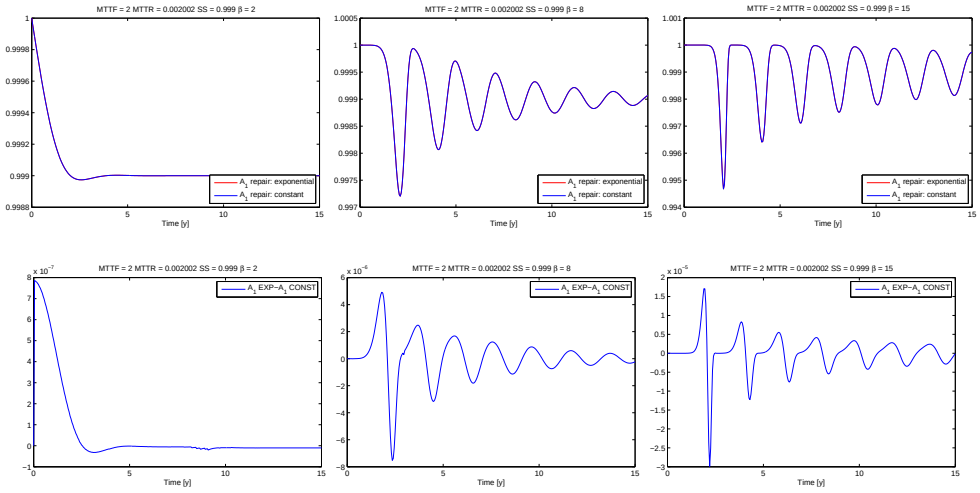


Figure 3.31: Comparison of the availability function for different repair times: steady state 0.999

1. The availability functions with high steady states are slightly affected by the distribution of down time.

2. For the realistic cases with small oscillations the difference between the availability functions with constant and exponential repair time ranges from 10^{-3} to

10^{-7}. Therefore if the accuracy of at least three digits is required the constant repair time may be used instead of the exponential one. If more digits are required then it is also possible to change the distributions of downtime but only for higher steady states.

3. For the high steady states (≈ 0.999) the difference between the availability functions with exponential and repair down times is negligible.

Chapter 4

Grace Period

This chapter presents a new approach for modeling the availability of a component in a situation when some failures are not visible.

4.1 Introduction

In practice we encounter the following situation. A component is put into operation at time t_0, when it fails it is repaired for some random time. If it takes a short time (less than x hours) to fix a failure, production can continue. If repair takes long, production has to cease after x hours. The x hours is called *grace period*. Similarly as before repair brings the component to the state "as good as new". The aim is to model production losses that are reflected only in "long" repairs. We introduce new state variable that represents the state (level) of production at time t

$$Y(t) = \begin{cases} 1 & \text{if production can continue at time } t \\ 0 & \text{otherwise} \end{cases}$$

As before we assume that up-times $(U_i, i = 1,2,\dots)$ are independent and identically distributed with distribution $F_U = P(U_i \leq t)$ and similarly down-times $(D_i, i = 1,2,\dots)$ are independent with common distribution $F_D = P(D_i \leq t)$. Previously we assumed that U_i and D_i are independent, however in this chapter we also allow those variables to be dependent. Example realization of the process $Y(t)$ is presented in Figure 4.1. Let

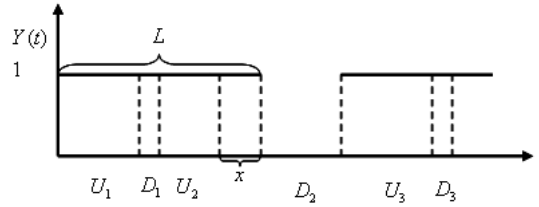


Figure 4.1: Example realization of process $Y(t)$

L_x denote the length of the first production time of a component with grace period x. Of

course the second production time will be different, however based on the assumption of independence and perfect repairs it will have the same distribution. Thus indexing of L_x is omitted to keep the notation clearer. It is easy to note that the new down-time is stochastically equal to $D_1 - x$. The aim of this chapter is to find the expected value and the distribution of the new down-time L_x.

Before the mathematical formulation of the problem is given we present some examples of situations that may occur in practice and are related to the grace period:

- Back-up electricity systems (sometimes called an Uninterruptible Power Supply (UPS)). If there is a failure of power supply in a computer center the processes on the computers can continue with the back-up electricity system for fixed number of hours. If restoration of power takes longer than the time the back-up system caters for the production stops and failure is visible. The maximal possible working time of the back-up system is the grace period.

- Flaring Allowance. In the oil industry flaring is used for burning off unwanted gas and liquids released by pressure relief valves during unplanned over-pressuring of plant equipment. Its primary purpose is to act as a safety device to protect vessels or pipes from over-pressuring. The allowed flaring time is restricted by the government regulations and varies from country to country. Therefore, when flaring occurs (due to the failure of some equipment, for instance a compression unit in a small side stream) the main production may still continue as long as the flaring allowance is not exceeded. The flaring allowance time is the grace period.

- Separation. One of the first stages of raw gas processing is water and condensate removal. Those substances have to be separated in order to process the condensate later on in a refinery. If separation of water and condensate is not possible due to failure of a separation unit, the substance can be send to an off-spec tank. The task of the tank is to keep the unseparated substance until the failure of the separation unit is fixed and there is some spare capacity to process additional substance from the tank. Production has to stop when during the repair of the separation unit the off-spec tank becomes full. The grace period in this case is the maximal time that the tank can take the substance before it is full. Note that if the failure of the separation unit occurs when the tank is not completely empty the grace period will be different and it should be established based on the capacity of the tank and its present fill.

From the practical point of view it is very important to model the availability of production. Note that in these situations the availability of a component is different than the availability of production because the component may be down and production may still continue. Therefore there is a need to develop mathematical model that allows to calculate the availability of production taking into account the grace period.

We use the same notation for up-times and down-times as in previous chapters.

Let $\{U_i, i = 1,2,\dots\}$ be the sequence of independent and identically distributed lifetimes having common distribution F_U, similarly let $\{D_i, i = 1,2,\dots\}$ be the sequence of independent and identically distributed repair times having common distribution F_D.

Let S_n denote the partial sum

$$S_n = \sum_{i=1}^{n} (U_i + D_i)$$

Then L_x is the sum of up-times and down-times for which consecutive repair times are smaller than x. During the first repair time that is greater than x the production can still continue for x time units therefore x has to be included in L_x as well. Hence we have that

$$L_x = S_{n-1} + U_n + x \quad \Leftrightarrow \quad D_1 \leq x, D_2 \leq x, \ldots, D_{n-1} \leq x, D_n > x \tag{4.1}$$

Furthermore L_x can be written as

$$L_x = \sum_{n=1}^{\infty} (S_{n-1} + U_n + x)\, \mathbf{1}_{\{D_1 \leq x, \ldots, D_{n-1} \leq x, D_n > x\}} \tag{4.2}$$

and since D_i are independent

$$L_x = \sum_{n=1}^{\infty} (S_{n-1} + U_n + x)\, \mathbf{1}_{\{D_1 \leq x\}} \cdot \cdots \cdot \mathbf{1}_{\{D_{n-1} \leq x\}} \mathbf{1}_{\{D_n > x\}} \tag{4.3}$$

4.2 Independent case

4.2.1 Expected value

Using representation (4.3) the expected value of L can be written as

$$\begin{aligned} E[L_x] &= \sum_{n=1}^{\infty} \left(E\left[(S_{n-1} + U_n + x)\, \mathbf{1}_{\{D_1 \leq x\}} \cdot \cdots \cdot \mathbf{1}_{\{D_{n-1} \leq x\}} \mathbf{1}_{\{D_n > x\}} \right] \right) \\ &= \sum_{n=1}^{\infty} \left(E\left[\sum_{i=1}^{n-1} (D_i + U_i)\, \mathbf{1}_{\{D_1 \leq x\}} \ldots \mathbf{1}_{\{D_{n-1} \leq x\}} \mathbf{1}_{\{D_n > x\}} \right] + E\left[(U_n + x)\, \mathbf{1}_{\{D_1 \leq x\}} \ldots \mathbf{1}_{\{D_{n-1} \leq x\}} \mathbf{1}_{\{D_n > x\}} \right] \right) \end{aligned} \tag{4.4}$$

Due to independence of D_i's and independence of U_i and D_i first expectation in above expression is

$$\sum_{i=1}^{n-1} \left(E\left[D_i \mathbf{1}_{\{D_i \leq x\}} \right] E\left[\prod_{j=1 j \neq i}^{n-1} \mathbf{1}_{\{D_j \leq x\}} \right] \cdot E\left[\mathbf{1}_{\{D_n > x\}} \right] + E[U_i] E\left[\prod_{j=1}^{n-1} \mathbf{1}_{\{D_j \leq x\}} \right] \cdot E\left[\mathbf{1}_{\{D_n > x\}} \right] \right)$$

Similarly the second expectation in (4.4) can be written as

$$(E[U_n] + x) E\left[\prod_{j=1}^{n-1} \mathbf{1}_{\{D_j \leq x\}} \right] \cdot E\left[\mathbf{1}_{\{D_n \geq x\}} \right]$$

Since the values $E[\mathbf{1}_{\{D_i \leq x\}}] = P(D_i \leq x)$ and $E[\mathbf{1}_{\{D_i > x\}}] = P(D_i > x)$ are independent on i we have that the expected value of L equals to

$$\begin{aligned} E[L_x] = \sum_{n=1}^{\infty} \big[&(n-1) \left(E\left[D_1 \mathbf{1}_{\{D_1 \leq x\}} \right] P(D_1 \leq x)^{n-2} P(D_1 > x) + E[U_n] P(D_1 \leq x)^{n-1} P(D_1 > x) \right) \\ &+ (E[U_n] + x)\, P(D_1 \leq x)^{n-1} P(D_1 > x) \big] \end{aligned}$$

and after rearranging and using the fact that $E[U_n]$ is independent on n

$$E[L_x] = \left[E\left[D_1\mathbf{1}_{\{D_1\leq x\}}\right]P(D_1>x)+E[U_1]P(D_1\leq x)P(D_1>x)\right]\sum_{n=1}^{\infty}(n-1)P(D_1\leq x)^{n-2}$$
$$+(E[U_1]+x)P(D_1>x)\sum_{n=1}^{\infty}P(D_1\leq x)^{n-1}$$

Since $|P(D_1\leq x)|<1$ both series in above expression are absolutely convergent, hence

$$\sum_{n=1}^{\infty}P(D_1\leq x)^{n-1}=\frac{1}{1-P(D_1\leq x)}$$

and

$$\sum_{n=1}^{\infty}(n-1)P(D_1\leq x)^{n-2}=\frac{d}{dP(D_1\leq x)}\left(\sum_{n=1}^{\infty}P(D_1\leq x)^{n-1}\right)=\frac{1}{[1-P(D_1\leq x)]^2}$$

Therefore

$$E[L_x] = \left(P(D_1>x)E\left[D_1\mathbf{1}_{\{D_1\leq x\}}\right]+E[U_1]P(D_1\leq x)P(D_1>x)\right)\frac{1}{(1-P(D_1\leq x))^2}$$
$$+(E[U_1]+x)P(D_1>x)\frac{1}{1-P(D_1\leq x)}$$

Using the fact that $P(D_1>x)=1-P(D_1\leq x)$ above expression may be simplified to

$$E[L_x]=\frac{E\left[D_1\mathbf{1}_{\{D_1\leq x\}}\right]+E[U_1]P(D_1\leq x)+(E[U_1]+x)P(D_1>x)}{1-P(D_1\leq x)} \tag{4.5}$$

where

$$E\left[D_1\mathbf{1}_{\{D_1\leq x\}}\right]=\int_0^x z dF_D(z). \tag{4.6}$$

Note that if there is no grace period i.e. $x=0$ then

$$E[L_0]=E[T_1]$$

also if $x\to\infty$ then

$$\lim_{x\to\infty}E[L_x]=\lim_{x\to\infty}\left(\frac{2E[D_1]}{1-P(D_1\leq x)}+E[D_1]+x\right)=\infty$$

as expected. In principle formula (4.5) allows to compute the expected value of production time including grace period for any choice of distributions of F_U and F_D that have finite first moment and support on $[0,\infty]$ (other distributions are not interesting in reliability modeling).
We copmare formula (4.5) with a simulation. As a test case we choose exponential distributions for both up-time and down-time, i.e. $F_U(t)=1-e^{-\lambda t}$ and $F_D(t)=1-e^{-\mu t}$. Then $E(U_1)=1/\lambda$ and $E(D_1)=1/\mu$. The integral in (4.6) equals to

$$E\left[D_1\mathbf{1}_{\{D_1<x\}}\right]=\int_0^x \mu z e^{-uz}dz=\frac{1-e^{-\mu x}-e^{-\mu x}x\mu}{u}$$

Substituting in (4.5) we have that for exponential-exponential case the expected production time including grace period x is

$$E[L_x] = \frac{\lambda+\mu-\lambda e^{-\mu x}}{\lambda\mu e^{-\mu x}} = \frac{\lambda+\mu}{\lambda\mu}e^{\mu\ x} - \frac{1}{\mu} \tag{4.7}$$

The simulation is based on random sampling of the production time L_x and comparing mean from the sample with the formula (4.5). The following parameters were used for the simulation: $\lambda = 0.2$, $\mu = 1$, $x = 1$, and hence $E[T_1] = 5$. Using formula (4.5) we obtain that $E[L] = 15.3097$. Figure 4.2 presents the estimated mean from samples with different size. We observe that results obtained from simulation agree with the theoretical value. Note that due to the grace period of length 1 mean production time

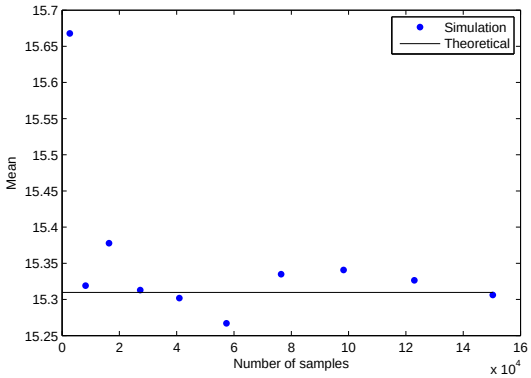


Figure 4.2: Comparison of the formula for $E[L_x]$ with simulation. Independent, exponential case.

has been extended from 5 to 15.3097. From the practical point of view it is important to know how the the extension of the grace period influences the mean production time. In this case the expected production time grows exponentially with respect to the grace period, see (4.7). In order to investigate this relationship let us consider an example with more realistic parameters. Let MTTF=5 years and MTTR=3 days. This corresponds to $\lambda = 1/5$, $\mu = 8766/(24\cdot 3)$. Figure 4.3 illustrates the relationship between the grace period and the expected production time for these parameters. This example shows that it is very important to introduce grace periods and were possible to extend them. For instance in this case introduction of the grace period of length 10 hours leads to the extension of the expected production time of almost 1 year!

4.2.2 The Distribution

As expected the distribution of L_x is much more difficult to obtain as compared to expected value. However it is not impossible. In this section we use technique of Laplace transforms that allows us to find the transform of the probability density function. Let $f_{L_x}(t)$ denotes the probability density function of the variable L_x and let $f^*_{L_x}(s)$ denotes its Laplace transform.

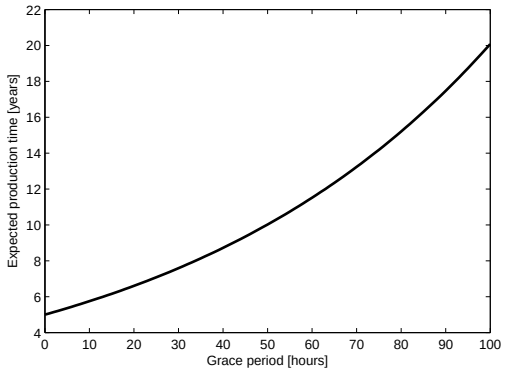


Figure 4.3: Relationship between the grace period and production time. Independent, exponential case.

Then we can write

$$f^*_{L_x}(s) = E[e^{-sL_x}] = \sum_{n=1}^{\infty} E[e^{-s(S_{n-1}+U_n+x)}; D_1 \leq x, \ldots, D_{n-1} \leq x, D_n > x] \tag{4.8}$$

$$= \sum_{n=1}^{\infty} E\left[\exp\left(-s\sum_{i=1}^{n} U_i - s\sum_{j=1}^{n-1} D_j - sx\right); D_1 \leq x, \ldots, D_{n-1} \leq x, D_n > x\right] \tag{4.9}$$

by independence of U_i and D_i we have that

$$= e^{-sx}\sum_{n=1}^{\infty} E\left[e^{-sU_1}\right]^n E\left[\exp\left(-s\sum_{j=1}^{n-1} D_j\right); D_1 \leq x, \ldots, D_{n-1} \leq x, D_n > x\right]$$

$$= e^{-sx}\sum_{n=1}^{\infty} E\left[e^{-sU_1}\right]^n E\left[e^{-sD_1}\cdot \mathbf{1}_{\{D_1\leq x\}}\right]^{n-1} P(D_1 > x)$$

$$= e^{-sx}E\left[e^{-sU_1}\right]P(D_1 > x)\sum_{n=1}^{\infty}\left(E\left[e^{-sU_1}\right]E\left[e^{-sD_1}\cdot \mathbf{1}_{\{D_1\leq x\}}\right]\right)^{n-1} \tag{4.10}$$

where in second line we used the independence of D_i's. For any positive random variable X and real $s > 0$ it holds that $|e^{-sX}| < 1$, this implies that $E[|e^{-sX}|] < 1$ from which it follows that $|E[e^{-sX}]| < 1$. Therefore the series in expression (4.10) is convergent and

$$\sum_{n=1}^{\infty}\left(E\left[e^{-sU_1}\right]E\left[e^{-sD_1}\cdot \mathbf{1}_{\{D_1\leq x\}}\right]\right)^{n-1} = \frac{1}{1 - E\left[e^{-sU_1}\right]E\left[e^{-sD_1}\cdot \mathbf{1}_{\{D_1\leq x\}}\right]}$$

Then the Laplace transform of the density function $f_L(t)$ becomes

$$f^*_{L_x}(s) = \frac{e^{-sx}P(D_1 > x)E\left[e^{-sU_1}\right]}{1 - E\left[e^{-sU_1}\right]E\left[e^{-sD_1}\cdot \mathbf{1}_{\{D_1\leq x\}}\right]} \tag{4.11}$$

or

$$f^*_{L_x}(s) = \frac{e^{-sx}P(D_1 > x)f^*_U(s)}{1 - f^*_U(s)E\left[e^{-sD_1}\cdot \mathbf{1}_{\{D_1\leq x\}}\right]} \tag{4.12}$$

where $f_U^*(s)$ denotes the Laplace transform of the up-time density. Expression (3.59), at least in principle, may be inverted for any choice of up- and down-time distributions and $f_L(t)$ may be found.
Assume again that that U_i and D_i have both exponential distribution with parameters λ and μ respectively. Then it is easy to show that

$$f_U^*(s) = E(e^{-sU_1}) = \frac{\lambda}{\lambda + s}$$

and

$$E\left[e^{-sD_1} \cdot \mathbf{1}_{\{D_1 \leq x\}}\right] = \int_0^x e^{-sz} f_D(z)dz = \mu \int_0^x e^{-sz-\mu z}dz = \frac{\mu}{\mu+s}\left(1 - e^{-x(s+\mu)}\right)$$

Substituting in formula (3.59) and simplifying we obtain that the the Laplace transform of density function $f_{L_x}(t)$ for exponential case becomes

$$f_L^*(s) = e^{-x(s+\mu)} \frac{\lambda(\mu+s)}{\lambda s + \mu s + s^2 + \mu\lambda e^{-x(s+\mu)}} \tag{4.13}$$

Unfortunately it is not straightforward how to invert above expression because it is not a fraction of polynomials for which analytical expression may always be found. However again we choose to compare obtained results with a simulation. Using the same parameters as for the simulation in previous section we sampled random variable L_x 19200 times. Now the empirical estimation of the density from this sample is compared to the numerical inversion of expression (3.60). For numerical inversion of the Laplace transform we used Euler method, see e.g. [1] On Figure 4.4 we observe that results are coherent.

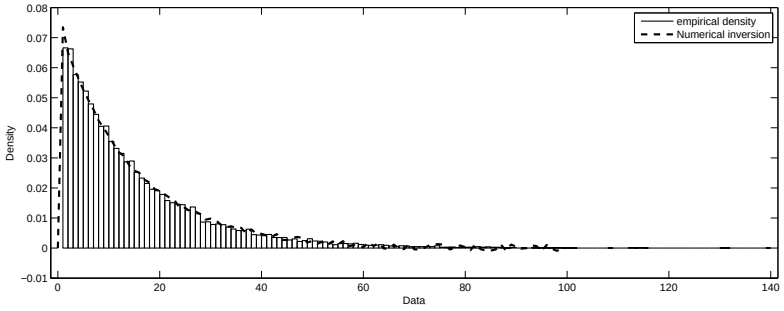


Figure 4.4: Comparison of the numerical inversion of equation (3.60) with empirical density.

Remark 4.2.1. Note that in general the support of $f_{L_x}(t)$ is $[x,\infty]$. This is due to the fact that production time will be at least equal to the grace period.

Remark 4.2.2. We want to investigate the behavior of $\lim_{t\to x+} f_{L_x}(t)$. It is known that for any continuous function f that have support on $[0,\infty]$ it holds $\lim_{t\to 0+} f(t) = \lim_{s\to\infty} sf^*(s)$, see e.g. [43]. Note that $f_{L_x}(t+x)$ have support on $[0,\infty]$ and

$$\lim_{t\to 0+} f_{L_x}(t+x) = \lim_{t\to x+} f_{L_x}(t)$$

The Laplace transform of $f_{L_x}(t+x)$ is

$$\begin{aligned}\mathcal{L}[f_{L_x}(t+x)] &= \int_0^\infty f_{L_x}(t+x)e^{-st}dt = \int_x^\infty f_{L_x}(z)e^{-s(z-x)}dz \\ &= e^{sx}\int_x^\infty f_{L_x}(z)e^{-sz}dz = e^{sx}\int_0^\infty f_{L_x}(z)e^{-sz}dz \\ &= e^{sx}f^*_{L_x}(s)\end{aligned}$$

Therefore in exponential case

$$\begin{aligned}\lim_{t\to x+} f_{L_x}(t) &= \lim_{s\to\infty} se^{sx}f^*_{L_x}(s) \\ &= \lim_{s\to\infty}\left(se^{-x\mu}\frac{\lambda(\mu+s)}{\lambda s+\mu s+s^2+\mu\lambda e^{-x(s+\mu)}}\right) \\ &= \lim_{s\to\infty}\left(e^{-x\mu}\frac{\lambda}{\frac{\lambda}{s}+1}\cdot\frac{1}{1-\frac{\lambda}{\lambda+s}\cdot\frac{\mu}{\mu+s}\left(1-e^{-x(s+\mu)}\right)}\right) \\ &= \lambda e^{-x\mu}\end{aligned}$$

Remark 4.2.3. Since $[\lambda e^{-x\mu}]^{-1} \neq E[L_x]$ from the possible candidates to the distribution of L_x we have to exclude exponential distribution (in fact shifted exponential) for which $E[X] = 1/\lim_{t\to 0} f_X(t)$.

Remark 4.2.4. Best candidate for the distribution. This remark has nothing to do with Laplace transforms. The candidate that fits the data the best (in maximum Likelihood sense) is Generalized Pareto distribution. Figure 4.5 presents the fit.

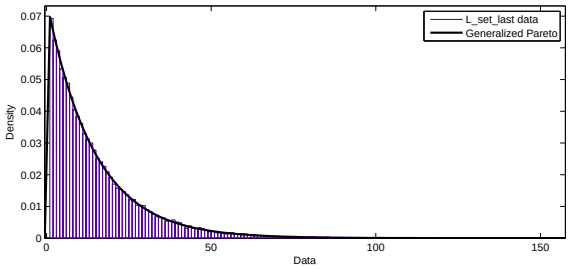


Figure 4.5: Generalized Pareto distribution as the best fit to data. Independent, exponential case.

Recall that Generalized Pareto distribution is a three parameter parameter family with the density

$$f(t) = \left(\frac{1}{\sigma}\right)\left(1+k\frac{t-\theta}{\sigma}\right)^{-1-\frac{1}{k}}$$

for $\theta < t$, when $k > 0$, or for $\theta < x < -\sigma/k$ when $k < 0$. Maximum likelihood estimation gave the following values of the parameters:

$$k = -0.00188266,\ \sigma = 14.2924$$

where the threshold parameter θ was specified to be equal to the grace period, in this case $\theta = 1$. It is interesting that in the limit $k \to 0$ the density becomes

$$f(t) = \left(\frac{1}{\sigma}\right) e^{-\frac{t-\theta}{\sigma}}$$

which is PDF of shifted exponentially distributed random variable. However as it was noticed in Remark 4.2.3 an exponential distribution is not allowed.

4.3 Dependent case

In this section we loose the assumption about the independence of up- and down-times in a cycle and allow U_i and D_i to be dependent. This assumption is far more realistic. First remark about the dependence structure between U_i and D_i is that they are expected to be positively correlated. Equipment that that has long time to failure requires long repair times. Another obvious relation is that usually lifetime is much bigger than repair time. One of the main aims of the equipment designers is to provide highly reliable components that can be repaired or replaced in a very short time as compared to the expected time to failure. These kind of dependencies may be, at least partially, modeled by suitable choice of the parameters to the distributions. However another example that occurs very often in practice cannot be treated by in that way. Consider a unit, say a pump in a plant, that operates 24 hours per day. The plant operates mainly during the day. When the pump fails during night shift the repair time may be much longer than during the normal working hours. From the mathematical point of view dependent case does not introduce any additional difficulties with deriving the formulas for the expected production time and its distribution.

4.3.1 Expected Value

Starting from representation (4.4) we have that expected production time with grace period x is

$$E[L_x] = \sum_{n=1}^{\infty} \left(E\left[\sum_{i=1}^{n-1} (D_i + U_i)\, \mathbf{1}_{\{D_1 \le x\}} \dots \mathbf{1}_{\{D_{n-1} \le x\}} \mathbf{1}_{\{D_n > x\}}\right] + E\left[(U_n + x)\, \mathbf{1}_{\{D_1 \le x\}} \dots \mathbf{1}_{\{D_{n-1} \le x\}} \mathbf{1}_{\{D_n > x\}}\right] \right)$$

Due to independence of U_i's first expectation in above expression may be written as

$$\sum_{i=1}^{n-1} \left(\left(E[D_i \mathbf{1}_{\{D_i \le x\}}] + E[U_i \mathbf{1}_{\{D_i \le x\}}]\right) \prod_{\substack{j=1 \\ j \ne i}}^{n-1} E[\mathbf{1}_{\{D_j \le x\}}] E[\mathbf{1}_{\{D_n > x\}}] \right)$$

Similarly the second expectation is

$$\left(E[U_n \mathbf{1}_{\{D_n > x\}}] + xE[\mathbf{1}_{\{D_n > x\}}]\right) \mathbf{1}_{\{D_1 \le x\}} \dots \mathbf{1}_{\{D_{n-1} \le x\}}$$

Using again the fact that U_i's are identically distributed we have that the expected value of L_x can be expressed as

$$\begin{aligned} E[L_x] = \sum_{n=1}^{\infty} \big[& (n-1) \left(E[D_1 \mathbf{1}_{\{D_1 \le x\}}] + E[U_1 \mathbf{1}_{\{D_1 \le x\}}]\right) P(D_1 \le x)^{n-2} P(D_1 > x) \\ & + \left(E[U_n \mathbf{1}_{\{D_n > x\}}] + xE[\mathbf{1}_{\{D_n > x\}}]\right) P(D_1 \le x)^{n-1} \big] \end{aligned}$$

We recognize the same infinite sums as in independent case, thus after doing some algebra we obtain that the expected production time with grace period x in dependent case is

$$E[L_x] = \frac{E[D_1\mathbf{1}_{\{D_1\le x\}}] + E[U_1\mathbf{1}_{\{D_1\le x\}}] + E[U_1\mathbf{1}_{\{D_1>x\}}] + xP(D_1>x)}{1-P(D_1<x)} \tag{4.14}$$

Note that if U_i and D_i, $i = 1,2,\ldots$ are independent then

$$E[L_x] = \frac{E[D_1\mathbf{1}_{\{D_1\le x\}}] + E[U_1]P(D_1\le x) + E[U_1]P(D_1>x) + xP(D_1>x)}{1-P(D_1<x)}$$

and we obtain formula (4.5).

4.3.2 The Distribution

Starting from representation (4.8) we have that the Laplace transform of L_x is

$$f^*_{L_x}(s) = \sum_{n=1}^{\infty} E\left[\exp\left(-s\sum_{i=1}^{n-1}(U_i+D_i) - sU_n - sx\right); D_1\le x,\ldots,D_{n-1}\le x, D_n>x\right]$$

Due to independence of U_i's above expression may be written as

$$\begin{aligned} f^*_{L_x}(s) &= e^{-sx}\sum_{n=1}^{\infty} E\left[\exp\left(-s\sum_{i=1}^{n-1}(U_i+D_i)\right)\mathbf{1}_{\{D_1\le x\}}\ldots\mathbf{1}_{\{D_{n-1}\le x\}}\right]E[e^{-sU_n}\mathbf{1}_{\{D_n>x\}}] \\ &= e^{-sx}\sum_{n=1}^{\infty} E\left[\exp(-s(U_1+D_1))\mathbf{1}_{\{D_1\le x\}}\right]^{n-1}E[e^{-sU_n}\mathbf{1}_{\{D_n>x\}}] \end{aligned}$$

Using the same argument about the convergance as in independent case we obtain that the Laplace transform of f_{L_x} is

$$f^*_{L_x}(s) = \frac{e^{-sx}E[e^{-sU_n}\mathbf{1}_{\{D_n>x\}}]}{1-E[e^{-s(U_1+D_1)}\mathbf{1}_{\{D_n\le x\}}]} \tag{4.15}$$

which in case when U_i and D_i, $i = 1,2,\ldots$ are independent reduces to the formula (3.59) for independent case. The expected value in the denominator, for non-negative random variables with the support $[0,\infty]$, is

$$E[e^{-s(U_1+D_1)}\mathbf{1}_{\{D_n\le x\}}] = \int_0^x\int_0^{\infty} f_{UD}(u,v)\,du\,dv$$

where $f_{UD}(u,v)$ denotes the joint probability density function of variables U_i and D_i.

Remark 4.3.1. Particle Counter Models. A counter is a device for detecting and registering radioactive particles. In general counters are imperfect and incapable of detecting all particles. After a particle is registered a counter must renew itself in preparation for next arrival. This time is called locked time. Two important classes of counting devices are two Type I and Type II counters. *Counters of Type I*: an arriving particle which finds the counter free gets registered and locks it for a time of length τ. Arrivals during locked period have no effect. *Counters of Type II*: an arrival particle during

the locked period does not get registered, but it extends the locked period so that the counter remains locked until τ units from that arrival. Length of locking period τ is assumed to be random. More detailed description of these counter models may be found in [**?**, 12]. Some possible deviations from such models are presented in [37]. A possible interpretation of the grace period in terms of those models is difficult. The closest is Type II counter model where the length of locking time depends on the number of particles arrived during the locked period and their locking times as it is in case with the grace period where production time depends on the number and length of up-times. However based on the literature research in this area we claim that no similar model as the grace period was developed.

Chapter 5

Conclusions

The theory of renewal processes has proven to be very useful in the context of applicability to the reliability modeling. Although the theory is rather simple and much research has been done in the last century there is still an open space for new models that can capture the reality more efficiently.

This thesis presents a complete framework for modeling the performance of a component based on its life and repair time distributions by using so called – availability functions. The advantage of the proposed approach is that it can be used to model the availability of a component taking into account its present state. The four functions developed for this reason are:

- $A_1(t)$ the component is new at time t_0
- $A_0(t)$ the repair of the component has just started at t_0
- $A_1^D(t)$ the component is functioning for x time units at time t_0
- $A_0^D(t)$ the component is in repair for y time units at time t_0

Additionally, the function $A_0^D(t)$ can be applied to model the availability of a component including maintenance schedules. This characterization is complete and gives a possibility to model any situation that may arise in practice. The examples given in Section show that there is a significant difference between $A_1(t)$ and $A_1(t)^D$, and their interchangeable usage is not suggested. A common practice in modeling the availability of a component that is x year-old, is to shift the function $A_1(t)$ by x years and use it instead of $A_1^D(t)$. However, this approach is not suggested, especially in case when the components are quite old since then the availability may drop much lower than in case when the method with shifting is used (see Figure 3.26(b)). The difference between the functions $A_0(t)$ and $A_0^D(t)$ is less significant and sometimes shifted function $A_0(t)$ may be used instead of $A_0^D(t)$, but this will depend on the repair time distribution. In general, however, the all four functions are used to model different situations and none of them should be omitted in the complete availability study. The state diagram presented in Appendix E illustrates the order in which those functions should be computed.

From the practical point of view it was relevant to find a method that is able to compute those functions with high accuracy and also within a reasonable time. This requirement has been met by using a combination of the adaptive quadrature for computing convolution together with TS method for numerical approximation of the renewal-type integral equation. The main criterion of the selection was the overall performance on many test problems that may arise in practice (not two or three examples like it is commonly practiced in literature). The accuracy of the approximation depends on two factors: error for computing convolution and the choice of the step-size used in the algorithm. Results show that in most of the cases with functions A_1, A_1^D the accuracy for convolution of order $10^{-7} - 10^{-9}$ is sufficient. For functions that starts from zero A_0, A_0^D the error in convolution can be of the same order but the step size for the main algorithm needs to be much smaller than in case of functions A_1, A_1^D. The speed of calculations is mainly affected by the time spent for computing the convolution and one may search for an additional improvement in this direction.

One of the most important aspect from the computational point of view is the steady state detection, that is the time after which the function will remain in some specified distance from its asymptotic value. It is not trivial to find a good estimation of the steady state before performing computations. The numerical investigation of the extremes, however, is a reasonable choice. The research in this direction would be extremely important with regards to applications (the location of low and high levels of production is crucial in the availability studies).

This research also showed that methods with bounds are not very useful from the practical point of view since they require too many grid points to be accurate, what is unacceptable when the convolution has to be approximated. However, there is an exception from this rule. The bounds for the availability and renewal functions may be used for short term predictions, usually up to time t for which $t \approx MTTF + MTTR$. In this situations the bounds are tight, even with small number of grid points.

We also claim that the methods based on the density function, like Volterra integral equation, are less applicable for real time calculations since they are too slow. Nevertheless if the high accuracy is required and the speed is not important one can use for instance RK-4 method that is able to produce very accurate results for either availability and renewal functions.

The availability functions for Gamma renewal alternating process, that were derived for the testing purposes, may also be used in practice instead of Weibull-Exponential case. The main advantage of using Gamma process is the simplicity in specifying the parameters for the distributions. It is very intuitive and requires only to answer the question: how many times the life time is bigger than the repair time? When the data is missing and there is not much known about the failure and repair mechanisms, the answer to this question may be the only one way to characterize a component and predict its availability.

It was also shown that the grace periods have a very important impact on the production time. For instance in case of exponentially distributed up and down times the the productions grows exponentially with respect to the grace period. Nevertheless the research in this area, especially with regards to distributions of the new production time, is still needed.

Bibliography

[1] J. Abate and W. Whitt. A unified framework for numerically inverting laplace transforms. *Informs Journal in Computing*, 18(4), 2006.

[2] M. K. Agrawal and S. E. Elmaghraby. On computing the distribution function of the sum of independent random variables. *Computers & Operations Research*, 28, 2001.

[3] H. Ayhan, J. Limon-Robles, and M.A. Wortman. An approach for computing tight numerical bounds on renewal functions. *IEEE Transactions on Reliability*, 48(2), June 1999.

[4] R. E. Barlow and F. Proschan. *Mathematical Theory of Reliability*. John Wiley & Sons, 1965.

[5] D.J. Bartholomew. An approximate solution of the integral equation of renewal theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, 25(2), 1963.

[6] L. A. Baxter. Availability measures for a two-state system. *Journal of Applied Probability*, 18, 1981.

[7] L. A. Baxter. Some remarks on numerical convolution. *Commun. Statist.*, 10(3):281–288, 1981.

[8] L. A. Baxter, E. M. Scheuer, D. J. McConalogue, and W. R. Blischke. On the tabulation of the renewal functions. *Technometrics*, 24, 1982.

[9] S. Bilgen and Z. S. Deligonul. Weibull renewal equation solution. *Computing*, 1986.

[10] H. Carlsson. Remainder term estimates of the renewal function. *The Annals of Probability*, 11(1):143–157, 1983.

[11] G. L. Choudhury and W. Whitt. Probabilistic scaling dor the numerical inversion of nonprobability transforms. *Informs Journal on Computing*, 9:175–184, 1997.

[12] E. Cinlar. *Introduction to Stochastic Processes*. Prentice-Hall, Inc., 1975.

[13] R. Clroux and D. J. McConalogue. A numerical algorithm for recursively-defined convolution integrals involving distribution functions. *Management Science*, 22(10), 1976.

[14] A.G. Constantine and N.I. Robinson. The weibull renewal function for moderate to large arguments. *Elsevier Science Publishers B. V.*, 24(1), 1997.

[15] D.R. Cox. *Renewal Theory*. Methuen's Monographs on Applied Probability and Statistics, 1962.

[16] L. Cui and M. Xie. Some normal approximations for renewal function of large weibull shape parameter. *Communication in Statistics, Simulation and Computation*, 32(1):1–16, 2003.

[17] L. M. Delves and J. L. Mohamed. *Computational methods for integral equations*. Cambridge University Press, 1985.

[18] Z. S. Doligonul. An approximate solution of the integral equation of the renewal therory. *Journal of Applied Probability*, 22:66–72, 1985.

[19] W. Feller. On the integral equation of renewal theory. *The Annals of Mathematical Statistics*, 12(3), September 1941.

[20] S.G. From. Some new approximations for the renewal function. *Statistical Computation and Simulation*, 30(1):113–128, 2001.

[21] W. Gander and W. Gautschi. Adaptive quadrature—revisited. *BIT Numerical Mathematics*, 40(1), 2000.

[22] A. Garg and J.R. Kalagnanam. Approximations for the renewal function. *IEEE Transactions on Reliability*, 47(1):March, 1998.

[23] W. Hackbusch. Fast and exact projected convolution for non-equidistant grids. *Computing*, 80(2):137–168, 2007.

[24] A. Hoyland and M. Rausand. *System Reliability Theory*. Wiley series in probability and mathematical statistics, 1994.

[25] X. Hu. Approximation of partial distirbution in renewal function calculation. *Computational Statistics & Data Analysis*, 50:1615–1624, 2006.

[26] S. Karlin and H. M. Taylor. *A First Course in Stochastic Processes*. Academic Press, 1975.

[27] Z.A. Lomnicki. A note on the weibull renewal process. *Biometrika*, 53(3/4), 1966.

[28] G. Lorden. On excess over the boundary. *Annals of Mathematical Statistics*, 41:520–527, 1970.

[29] K. T. Marshall. Linear bounds on the renewal function. *SIAM Journal on Applied Mathematics, Vol. 24, No. 2 (Mar., 1973), pp. 245-250*, 24(2):245–250, 1973.

[30] D. J. McConalogue. Numerical treatement of convolution integrals invoving distibutions with densitites having singularites at the origin. *Communications in Statistics*, 1981.

[31] S. Mercier. Discrete random bounds for general random variables and applications to reliability. *European Journal of Operational Research*, 127(1):378–405, February 2007. available at http://ideas.repec.org/a/eee/ejores/v177y2007i1p378-405.html.

[32] G. Muraleedharan and A. D. Rao et al. Modified weibull distribution for maximum and signficant wave height simulation and prediction. *Coastal Engineering*, 54, 2007.

[33] V. K. Murthy. *The General Point Process: Applications to Structural Fatigue*. Addison-Wesley, Reading, Mass., 1974.

[34] F. Oberhettinger and L. Badii. *Tables of Laplace Transforms*. Springer-Verlag, 1973.

[35] E. Parzen. *Stochastic Processes*. Holden-Day Series in Probability and Statistics, 1962.

[36] T. Pham-Gia and N. Turkkan. System availability in a gamma alternating renewal process. *Naval Research Logistics (NRL)*, 46(7):822–844, 1999.

[37] R. Pyke. On renewal processes related to type i and type ii counter models. *The Annals of Mathematical Statistics*, 29(3):737–754, 1958.

[38] L. Ran, L. Cui, and M. Xie. Some analytical and numerical bounds on the renewal function. *Communications in Statistics, Theory and Methods*, 35:1815–1827, 2006.

[39] E. Smeitink and R. Dekker. A simple approximation to the renewal function. *IEEE transactions on reliability*, 39(1), April 1990.

[40] W. L. Smith. Asymptotic renewal theorems. proc. roy. soc. edin. a, 64, 9-48. *Proceedings of the Royal Society of Edinburgh, Section A (Mathematical and Physical Sciences)*, 64:9–48, 1954.

[41] W.L. Smith. Renewal theory and its ramifications. *Journal of the Royal Statistical Society*, 20(2), 1958.

[42] W.L. Smith and M.R. Leadbetter. On the renewal function for the weibull distribution. *Technometrics*, 5(3), August 1963.

[43] M. R. Spiegel. *Theory and Problems of Laplace Transforms*. Schaum's Outline Series, 1965.

[44] H. C. Tijms. *A first Course in Stochastic Models*. New York, Wiley, 2003.

[45] F. A. Tillman, W. Kuo, R. F. Nassar, and C. L. Hwang. Numerical evaluation of instantaneous availability. *IEEE Transactions on Reliability*, R-32, 1983.

[46] M. Tortorella. Closed newton-cotes quadrature rules for stieltjes integrals and numerical convolution of life distributions. *SIAM J. Sci. Stat. Comput.*, 11(4):732–748, 1990.

[47] M. Tortorella. Numerical solutions of renewal-type integral equations. *INFORMS J. on Computing*, 17(1):66–74, 2005.

[48] M. Xie. On the solution of renewal-type integral equations. *Communications in Statistics Computation & Simulation*, 47(1):66–72, 1989.

Appendix A

The Distributions

Below we present the characteristics of the distributions that are used in this report. Let X be a random variable. We specify the distribution of X by either density or cumulative distribution function.

1. Exponential Distribution. The probability density function of an exponential distribution has the form

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

where $\lambda > 0$ is a parameter of the distribution, often called the rate parameter. The distribution is supported on the interval $[0, \infty)$. The cumulative distribution function is given by

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

2. Weibull Distribution. The probability density function is:

$$f(x) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha} \right)^{\beta - 1} e^{-(x/\alpha)^{\beta}}$$

for $x > 0$ and $f(x) = 0$ for $x \leq 0$, where $\beta > 0$ is the shape parameter and $\alpha > 0$ is the scale parameter of the distribution. The cumulative distribution function for is

$$F(x) = 1 - e^{-(x/\alpha)^{\beta}}$$

for $x \geq 0$, and $F(x) = 0$ for $x < 0$. The characteristic function of Weibull distribution is (see [32])

$$\Phi(s) = 1 + \sum_{r=0}^{\infty} \frac{(is\alpha)^{r+1}}{r!\beta} \cdot \Gamma\left(\frac{r+1}{\beta} \right)$$

The Laplace transform of Weibull distribution is (see **??**)

$$f^*(s) = \sum_{r=0}^{\infty} \frac{(-1)^r (s\alpha)^r}{r!} \cdot \Gamma\left(\frac{r}{\beta} + 1 \right)$$

3. Lognormal Distribution. The lognormal distribution has the probability density function
$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$
where $-\infty < \mu < \infty$ and $\sigma > 0$. Expectation and variance are
$$E(X) = e^{\mu+\sigma^2/2} \quad \mathrm{Var}(X) = (e^{\sigma^2}-1)e^{2\mu+\sigma^2}.$$
We remark that there is no closed form for the cumulative distribution function for lognormally distributed random variable and its CDF has to be approximated (the most common approximation is in terms of error function).

4. Constant Distribution. The CDF of this discrete distribution is
$$F(x) = \begin{cases} 1 & , x \geq \tau, \\ 0 & , x \leq \tau. \end{cases}$$

5. Gamma Distribution. That a random variable X is gamma-distributed with scale λ and shape k if its PDF can be expressed as
$$f(x) = \begin{cases} \frac{\lambda}{(k-1)!}(\lambda x)^{k-1}e^{-\lambda x} & , x \geq 0, \\ 0 & , x < 0. \end{cases}$$
Similarly as in lognormal case the CDF has to b approximated.

6. Remaining Life Time distribution. Of our interest is the distribution of the remaining life time of a component given that it is in operation for x time units. Let T be a random variable with distribution $F_T(t)$ and let $F_{T_x}(t)$ denotes the probability that the component will fail in time $t+x$ given that it was in operation for x time units. Then
$$1 - F_{T_x}(t) = P(T > t+x | T > x) = \frac{P(T > t+x, T > x)}{P(T > x)} = \frac{P(T > t+x)}{P(T > x)}.$$
therefore
$$F_{T_x}(t) = 1 - \frac{1 - F_T(t+x)}{1 - F_T(x)}.$$
In principle this distribution may be found for any choice of F_T. For instance, if F_T is Weibull distribution then
$$F_T(t) = 1 - e^{-\left(\frac{t}{\alpha}\right)^\beta}.$$
It follows that
$$\frac{P(T > t+x)}{P(T > x)} = \frac{e^{-\left(\frac{t+x}{\alpha}\right)^\beta}}{e^{-\left(\frac{x}{\alpha}\right)^\beta}} = \exp\left[-\left(\frac{t+x}{\alpha}\right)^\beta + \left(\frac{x}{\alpha}\right)^\beta\right]$$
thus
$$F_{T_x}(t) = 1 - \exp\left[-\left(\frac{t+x}{\alpha}\right)^\beta + \left(\frac{x}{\alpha}\right)^\beta\right]$$

Since $F_{T_x}(t)$ correpsonds to a continuous random variable it has a density. By simple differentiation we obtain that

$$f_{T_x}(t) = \left(\frac{t+x}{\alpha}\right)^{\beta} \beta e^{-\left(\frac{t+x}{\alpha}\right)^{\beta}+\left(\frac{x}{\alpha}\right)^{\beta}} (t+x)^{-1}$$

Appendix B

Laplace Transforms

Suppose that $f(x)$ is a function that is defined on the interval $(0,\infty)$. The Laplace transform $f^*(s)$ of the function $f(t)$ is defined by

$$f^*(s) = \int_0^\infty e^{-st} f(t)dt$$

where s is a real number. It is convenient sometimes to use the alternative notation

$$\mathcal{L}[f(t)] = f^*(s)$$

Not all functions have a Laplace transform, for instance if $f(t) = e^{t^2}$ the integral diverges for all values of s. When $f(t)$ is the probability density function of a nonnegative random variable X, the Laplace transform of $f(t)$ is seen to be equal the expected value of the random variable e^{-sX}:

$$E[e^{-sX}] = \int_0^\infty e^{-st} f(t)dt = f^*(s)$$

The function $f(t)$ us called the inverse Laplace transform of $f^*(s)$, and is written

$$f(t) = \mathcal{L}^{-1}[f^*(s)]$$

As an example, consider the exponential distribution $f(t) = \lambda e^{\lambda t}$. The corresponding Laplace transform is

$$\int_0^\infty \lambda e^{-\lambda t} e^{-st} = \frac{\lambda}{\lambda + s}$$

The Laplace transforms of over 300 functions may be found in [34]
Properties

1. $\mathcal{L}[f_1(t) + f_2(t)] = \mathcal{L}[f_1(t)] + \mathcal{L}[f_1(t)]$
2. $\mathcal{L}[af(t)] = a\mathcal{L}[f(t)]$
3. $\mathcal{L}[f(t-a)] = e^{-at}\mathcal{L}[f(t)]$

4. $\mathcal{L}[f'(t)] = s\mathcal{L}[f(t)] - f(0)$
5. $\mathcal{L}\left[\int_0^t f(u)du\right] = \mathcal{L}[f(t)]/s$
6. $\mathcal{L}\left[\int_0^t f_1(t-u)\, f_2(u)du\right] = \mathcal{L}[f_1(t)] \cdot \mathcal{L}[f_2(t)]$
7. Initial value theorem: $\lim_{s\to\infty} s f^*(s) = \lim_{t\to 0} f(t)$
8. Final value theorem: $\lim_{s\to 0} s f^*(s) = \lim_{t\to\infty} f(t)$
9. If $f(t)$ is a probability density function then $0 < f^*(s) \leq 1$ for all $s \geq 0$

Proofs of properties 1-6 are based on the properties of integration. Properties 7 and 8 are derived by using property 4 and the last property is a simple consequence of the property of an exponential function. All properties are not difficult to prove and may be found in standard textbooks.

Appendix C

Adaptive Simpson's Rule

This appendix describes the idea of the numerical integration by using adaptive grid.

Let $[a,b]$ be the interval of integration, assumed to be bounded, and let f be a real integrable function. We are interested in computing the integral

$$Q = \int_a^b f(x)dx. \tag{C.1}$$

The idea of the adaptive grid is as follows. Compute two approximations of Q using two different numerical integration methods to obtain Q_1 and Q_2. We assume that one approximation, say Q_1 is more accurate than the other. If the relative difference between Q_1 and Q_2 is smaller than some prescribed tolerance than one accepts Q_1 as the value of the integral. Otherwise the interval $[a,b]$ is divided in two equal parts $[a,m]$, $[m,b]$, where $m = (a+b)/2$, and two integrals

$$\int_a^m f(x)dx.$$

and

$$\int_m^b f(x)dx.$$

are computed independently. One now again computes recursively two approximations Q_1 and Q_2 for each integral and, if necessary, continues to subdivide the smaller intervals. This is the basic idea of the adaptive grid. For approximation of Q_1 and Q_2 we choose Simpson's rule with step size $h/2$ and h respectively (we expect more accurate approximation with smaller step size). The advantage of choosing the same quadrature is that for every iteration we can compute more accurate approximation to Q using one iteration of so called Romberg extrapolation. More precisely, we know that for Simpson's rule the relation between the integral and its approximation is of the form

$$Q = Q(h) + Ah^4$$

where A is some constant independent on h and $Q(h)$ denotes approximation of Q by Simpson's rule with step size h. Similarly for $Q(h/2)$ we have

$$Q = Q(h/2) + A(h/2)^4.$$

We can find constant A from the first equation and substitute in the second one. Therefore we obtain that

$$Q = \frac{16Q(h/2) - Q(h)}{15}.$$

The error in approximation now has order of h^6 instead of initial h^4.
For approximations of Q_1 and Q_2 we decide to choose the following Simpson's rule

$$Q_1 = \frac{h}{6}\left(f(a) + 4f(m) + f(b)\right)$$

and

$$Q_2 = \frac{h}{12}\left(f(a) + 4f(d) + 2f(m) + 2f(e) + f(b)\right)$$

where $d = (a+m)/2$ and $e = (m+c)/2$.

Stopping criteria

Using adaptive grids we have to be very careful in choosing stopping criterion. We suggest to use the following

1. Stop when the difference between two consecutive approximations is less than some prescribed tolerance.

2. The other problem we need to take care of is when an interval gets subdivided so small that it contains no interior machine-representable point. Therefore we terminate the recursion when the step size is smaller than than minimal possible step size or when $c \geq a$ or $c \leq b$.

Appendix D

Proof of Theorem 3.5.1

We prove Theorem 3.5.1 by using technique of Laplace transforms. Let us define

$$\phi(u) = \sum_{n=1}^{\infty} \frac{u^{nk-m}}{(nk-m)!}$$

The Laplace transform of ϕ is

$$\begin{aligned}\phi^*(s) &= \int_0^\infty e^{-su}\phi(u)du = \int_0^\infty e^{-su} \sum_{n=1}^{\infty} \frac{u^{nk-m}}{(nk-m)!}du \\ &= \sum_{n=1}^{\infty} \frac{1}{(nk-m)!} \int_0^\infty e^{-su}u^{nk-m}du\end{aligned}$$

where interchange of sum and integral is justified by absolute convergence of the series. It can be easily verified that by changing variables to $su = x$ the Laplace transform becomes

$$\phi^*(s) = \sum_{n=1}^{\infty} \frac{1}{(nk-m)!} \frac{1}{s^{nk-m+1}} \int_0^\infty e^{-x}x^{nk-m}du$$

Since for any integer z

$$\int_0^\infty e^{-x}x^{z-1}du = \Gamma(z) = (z-1)!$$

then

$$\begin{aligned}\phi^*(s) &= \sum_{n=1}^{\infty} \frac{1}{(nk-m)!} \frac{1}{s^{nk-m+1}}(nk-m)! = s^{m-1} \sum_{n=1}^{\infty} \left(\frac{1}{s^k}\right)^n \\ &= \frac{s^{m-1}}{s^k-1} \qquad \text{(D.1)}\end{aligned}$$

Since (D.1) is a fraction of two polynomials it can be inverted by the method of partial fractions. First we need to find roots s_r $r = 0,\ldots,k-1$ of the denominator s^k-1. It

can be easily verified that those roots are $s_r = \exp\left(\frac{2\pi i}{k}\right)^r$. Let us denote $\varepsilon = \exp\left(\frac{2\pi i}{k}\right)$. Therefore the denominator can be expressed as

$$s^k - 1 = (s-\varepsilon^0)(s-\varepsilon^1)\cdot\cdots\cdot(s-\varepsilon^{k-1})$$

Then the Laplace transform of ϕ can be written as

$$\frac{s^{m-1}}{s^k-1} = s^{m-1}\sum_{j=0}^{k-1}\frac{A_j}{s-\varepsilon^j} \tag{D.2}$$

Multiplying both sides of (D.2) by $s-\varepsilon^r$ gives

$$\frac{s^{m-1}}{s^k-1}(s-\varepsilon^r) = s^{m-1}\sum_{j=0,j\neq r}^{k-1}\frac{A_j}{s^k-1}(s-\varepsilon^r) + s^{m-1}A_r \tag{D.3}$$

From (D.3) it follows that coefficients A_r are

$$A_r = \lim_{s\to\varepsilon^r}\frac{s-\varepsilon^r}{s^k-1} \tag{D.4}$$

Above limit can be computed using d'Hôpital's rule

$$A_r = \lim_{s\to\varepsilon^r}\frac{1}{k(\varepsilon^r)^{k-1}} = \frac{\varepsilon^r}{k} \tag{D.5}$$

Therefore the Laplace transform of ϕ becomes

$$\phi^*(s) = \frac{1}{k}\sum_{r=0}^{k-1}\varepsilon^r\frac{s^{m-1}}{s-\varepsilon^r} \tag{D.6}$$

The fraction in above expression may be rewritten as follows

$$\begin{aligned}\frac{s^{m-1}}{s-\varepsilon^r} &= \frac{s^{m-1}-(\varepsilon^r)^{m-1}+(\varepsilon^r)^{m-1}}{s-\varepsilon^r}\\ &= \frac{(s-\varepsilon^r)\left(s^{m-2}+s^{m-3}\varepsilon^r+s^{m-4}(\varepsilon^r)^2+\cdots+s(\varepsilon^r)^{m-3}+(\varepsilon^r)^{m-2}\right)+(\varepsilon^r)^{m-1}}{s-\varepsilon^r}\\ &= \sum_{j=0}^{m-2}s^{m-2-j}(\varepsilon^r)^j + \frac{(\varepsilon^r)^{m-1}}{s-\varepsilon^r}\end{aligned}$$

Thus

$$\phi^*(s) = \frac{1}{k}\sum_{r=0}^{k-1}\varepsilon^r\sum_{j=0}^{m-2}s^{m-2-j}(\varepsilon^r)^j + \frac{1}{k}\sum_{r=0}^{k-1}\varepsilon^r\frac{(\varepsilon^r)^{m-1}}{s-\varepsilon^r} \tag{D.7}$$

First term in the above expression is zero

$$\frac{1}{k}\sum_{r=0}^{k-1}\varepsilon^r\sum_{j=0}^{m-2}s^{m-2-j}(\varepsilon^r)^j = \sum_{j=0}^{m-2}s^{m-2-j}\frac{1}{k}\sum_{r=0}^{k-1}\left(\varepsilon^{j+1}\right)^r = \frac{1-\varepsilon^{(j+1)k}}{1-\varepsilon^{j+1}} = 0 \tag{D.8}$$

Straightforward inversion of the second term in (D.7) completes the proof.

Appendix E

State diagram for calculations

This appendix presents the state diagram for computing the availability functions that may facilitate the implementation of the algorithms.

In practice one computes the availability function until it converges to the steady state (with some desired accuracy) and after that the asymptotic value for the function is used (no calculations are necessary). Before the steady state is reached there may be a planned maintenance scheduled at time t_{PM}, then the function is computed on $[0, t_{PM}]$ and at $t = t_{PM}$ the availability goes to zero. For $t > t_{PM}$ the function A_{PM} is used. Planned maintenance may be also scheduled after the steady state is reached this applies to all four availability functions. Therefore in order to compute the availability of a component including the planned maintenance and steady states we propose the following diagram for easier implementation of the possible situations with the availability functions.

Description of the states:

0 : Before the calculations

1 : Unit at start-up: function A_1

2 : Unit at start of repair: function A_0

3 : Unit in operation for x hours: function A_1^D

4 : Unit in repair for y hours: function A_0^D

5 : Planned maintenance just started: function A_{PM}

6 : Availability in steady-state: function A (constant)

This diagram can be explained as follows. At any given time a component may be in one of the five states: start-up, start of repair, in operation, during the repair or at the beginning of planned maintenance therefore from the state 0 it is possible to go to states 1,2,3,4,PM [1]. Let us first consider state 1. The calculations may be performed

[1] We do not consider the steady state as the transition from the state 0 since there is no need to do calculations for steady state

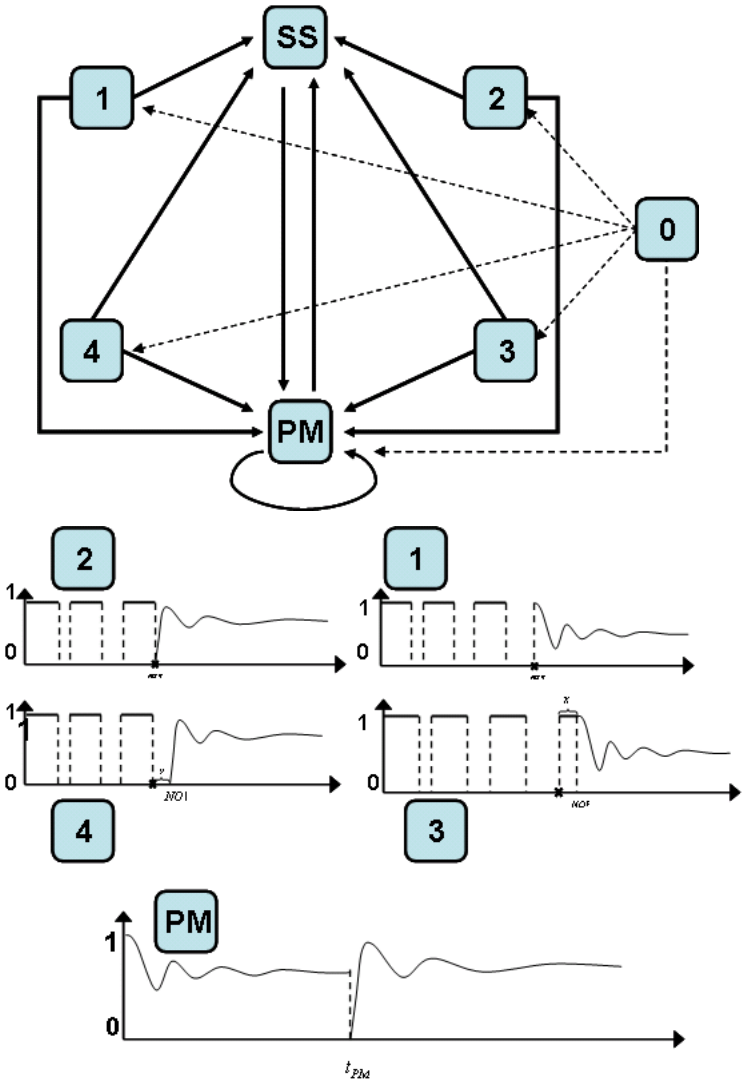


Figure E.1: State diagram for computing availability.

until the nearest planned maintenance time t_{PM} or until the steady state thus transition from state 1 to SS or PM. From SS it is possible to have another planned maintenance thus transition from SS to PM. The calculations may stay in state PM since there may be another maintenance scheduled, however it is also possible to reach the steady state thus also transition from PM to SS is possible. When there are no planned maintenances scheduled the calculations will eventually end up in the SS state. The transitions from states 2,3 and 4 can be described in a similar manner.