NTNU – Trondheim
Norwegian University of
Science and Technology

A Simulation-based Approach for Exploring the Degradation and Maintenance of a Single-unit System

Odd Eirik Farestveit

Master of Science in Physics and Mathematics
Submission date: June 2015
Supervisor: Bo Henry Lindqvist, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem description

- Give an introduction to stochastic processes relevant to degradation modeling.
- Demonstrate the use of such models for maintenance planning.
- Develop a simulation based algorithm for the exploration of the models, with emphasis on its theoretical basis.

Assignment given: January 16, 2015.

Supervisor: Bo Henry Lindqvist (NTNU).

Abstract

A model for degradation and maintenance of a single-item system is extensively studied in this thesis, using a simulation based approach. Stochastic processes are used as degradation models and the item is maintained according to a condition-based maintenance program. Necessary background theory on stochastic process and the field of reliability is presented. This theory is linked directly to different properties of the model. Amongst other things, the concepts of regenerative and semi-regenerative processes will be important, both for computational and simulation related purposes. A Monte Carlo based algorithm is developed to simulate and explore the model. Within this algorithm, bridge sampling for simulation of stochastic processes play an important role. Numerical experiments with the model and possible extensions of it, is presented towards the end of the thesis.

Samandrag

Denne oppgåva er ein omfattande simuleringsbasert studie av ein slitasje- og vedlikehaldsmodell for eit ein-komponent system. Stokastiske prosessar er nytta som slitasjemodellar og komponenten vert haldt vedlike gjennom eit tilstandsbasert vedlikehaldsprogram. Oppgåva gjennomgår naudsynt teori om både stokastiske prosessar og pålitelegheitsanalyse. Denne teorien relaterast deretter til ulike eigenskapar ved modellen. Mellom anna er konsepta regenerative og semi-regenerative prosessar viktige, på grunn av følgjene dei har for utrekningsmetodar og av meir simuleringstekniske årsakar. Vidare gjerast det greie for ein Monte Carlo-basert algoritme for å simulere og utforske modellen. Ein viktig del av denne algoritmen, er å kunne simulere ein stokastisk prosess mellom to gjevne observasjonar av prosessen. Mot slutten av oppgåva presenterast ei rekkje numeriske eksperiment med både modellen og moglege utvidingar av den.

Preface

This text is my Master's thesis and marks the end of my studies at NTNU's advanced engineering program in Applied Physics and Mathematics. It has been five great years and I would like to thank my classmates and all my friends in Trondheim for a wonderful time. Prior to my time in Trondheim, I attended a teacher program at Sogn og Fjordane Univeristy College (HiSF) in Sogndal and I would also like to thank my friends back there for a lot of good memories. This was where i discovered and got curious about mathematics and I'm very grateful to Anne Norstein and Terje Myklebust at HiSF, for instructing an excellent one year course in mathematics and didactics in the academic year 2009/2010.

A special thanks also go to professor Anne Barros at NTNU's department of Production and Quality Engineering, for providing some valuable litterature on semi-renegenerative processes.

Finally and most importantly I would like to say thank you to my excellent supervisor, professor Bo Lindqvist at NTNU's department of Mathematical Sciences, for his help both with my specialization project last autumn and with this thesis. Working with a Master's thesis is not always straight forward, and I'm therefore very grateful for the discussions we had, the advices he gave and his patience with me.

Odd Eirik Farestveit

Trondheim, June 2015.

Contents

1 Introduction

Throughout this text, we will present a model that describes the degradation of a generic item, along with a maintenance policy. By the term item, we mean a singe-unit system, performing some desired task, for example a production machine or a pipeline of flowing gas. This model was originally proposed in [11] and [9]. We will describe both mathematical properties and possible extensions of the model. By using Monte Carlo simulations in our numerical exploration of this model, we differ from the original papers, where a numerical optimization procedure is performed. An important cornerstone of the model, is that it uses a stochastic process to describe the time evolution of the item's degradation. In practice, this requires both the ability to measure the item's state and to handle a larger amount of data than we do in classical reliability analysis.

There are reasons to believe that the field of reliability analysis can change rapidly in the years to come. Amongst these reasons, the ongoing development on analysis of large data sets, often called big data analysis, is maybe of the most important one. For a nice and non-formal introduction to this topic, we recommend to read [13], where both different applications and types of data (and how they arrive) are reviewed. In this article the author also stresses the need for statisticians to learn and contribute to big data analysis, as this topic is placed in the intersection between the interests of computer scientists, business researchers and statisticians. An example of a popular application is for a business to understand how individual customers behave, and be able to tailor future advertisement directly to a particular consumer.

Other, and perhaps more important applications can be found in the field of medical statistics and reliability. Our context will be reliability, but the theory of these two fields often goes hand in hand. [16] gives an excellent introduction on big data reliability analysis. Their point of departure is that new technology, such as sensors and smart chips in systems, can collect an enormous amount for so-called system operating/environmental-data (SOE-data). By exploiting these data, one should be able to improve on a system's performance. The authors presents a wide variety in examples of methods and fields of application. Among these methods, degradation modeling and condition based maintenance are mentioned specifically.

Having been in contact with representatives from the reliability business from both inside and outside Norway, it is clear that most of the models used on real world problems today are much simpler, seen from a statistician's point of view. Instead the applied models often uses more qualitative information, together with classical reliability analysis, where estimates often are provided by reliability databases and manufacturers. However, there is an increasing trend of monitoring systems and an increasing interest of exploiting SOE-data. [15] is a publication from the large certification and classification society DNV GL, where both ongoing projects and possible opportunities in connection with large amount of data collection, is presented. It concerns more applications than just reliability, but we can note that collection of SOE-data in the oil and gas industry, and proactive maintenance of wind turbines are among the things listed. Many similar reports are also available from other large organisations and companies, concerning a wide range of businesses, also including health care and medical statistics.

All this shows that creation, understanding and implementation of more complex models in reliability and maintenance optimization is an area where a lot research are

being done, both in the academic world and the industry. In this text we will only consider one such model, but the big picture discussed above, should serve as a motivation of this study. The model that we will explore was actually proposed in 2002, but it might even be more relevant today.

We will now give an outline of the content in this thesis. Chapter 2 first introduces continuous state space Markov processes and stationary distributions. Then we give a quick introduction to Renewal theory, starting with the basics ending with the so-called Markov-Renewal process. We close this chapter by establishing a link from the Renewal theory to regenerative and semi-regenerative processes. In chapter 3, we present some background the maintenance theory and degradation processes. The Gamma-process and the Inverse Gaussian process are also defined and discussed with the purpose of being used as degradation models. The original degradation/maintenance-model, along with some extensions are presented in chapter 4. Here we also discuss properties of this model in light of the theory from chapter 2 and 3. Chapter 5 explains how a computer program for simulating the models can be implemented. The derivation of an algorithm for simulating passage times is a key result, which builds upon so-called bridge sampling algorithm for the underlying degradation process. Having described the implementation, we run a series of numerical experiments in chapter 6. The goal of the simulations is to explore how the systems behaves when we change the parameters or extend the model. Finally chapter 7 provides a discussion on the work being done and suggestions for future work.

2 Stochastic processes

2.1 Markov processes

Without giving a formal mathematical definition of stochastic processes, we follow [8] and refer to a stochastic process as a collection of random variables X_t, defined for all relevant values of t. In our context t is positive, denoting time and can be both discrete or continuous. All possible values for an individual X_t are called the state space. This text will be dealing with one-dimensional real stochastic processes, meaning that the state space at any time is a one-dimensional set $\mathcal{X} \subset \mathbb{R}$. It is natural to distinguish between four types of one-dimensional stochastic process, that is

- Discrete time, discrete state space.
- Discrete time, continuous state space.
- Continuous time, discrete state space.
- Continuous time, continuous state space.

Processes covered by us will may both have continuous or discrete time, but always continuous state space. The reader should therefore be aware that some of the equations in this section may have discrete versions. At time t, we define $\mathcal{H}_t$ to be the history of the process, meaning the collection of all points (s, X_s), $s \leq t$ that the process has passed through on its way up to time t, (t, X_t) included. An important class of stochastic processes are the Markov processes:

Definition 1. *A stochastic process $X = \{X_t, t \geq 0\}$ is said to be a Markov process if and only if (for any $s > 0$)*

$$\mathrm{E}\left[f\left(X_{t+s}\right)|\mathcal{H}_t\right] = \mathrm{E}\left[f\left(X_{t+s}\right)|X_t\right],$$

for any continuous function f.

That is, the distribution of future states only depends on the present state, not the past. If a stochastic process is a Markov process, it is sometimes just referred to as being Markovian. Consider now two finite time point s and t, where $0 \leq s < t$. By an increment of a stochastic process we mean the random varibale $X_t - X_s$, representing the change in process from time s to time t. Refering to [1], we say that a stochastic process has stationary increments if for any set $\mathcal{A} \subset \mathcal{X}$

$$P\left(X_t - X_s \in \mathcal{A}\right) = P\left(X_{t-s} - X_0 \in \mathcal{A}\right),$$

which means that the distribution of the increments only depends on the time difference $t - s$, not the time points s and t. The increments are said to be independent if for any collection of finite time points $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$, the random variables $X_{t_1} - X_0, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent. It can be shown that independent increments implies that a process is Markovian. In what follows we will go through a series of important concepts and properties regarding Markov processes. These characteristics are mostly a summary of discussions found in [8] and [10]. It should also be mentioned that we simplify the content a bit and some of the definitions and results are presented more formally in

these textbooks. Let $X = \{X_n, n \geq 0\}$ be a discrete time Markov process with continuous state space $\mathcal{X}$, and let ν is a probability distribution defined on the state space $\mathcal{X}$ and denote its density function by $\nu(x)$. For any real-valued function f on $\mathcal{X}$, the expectation of f with respect to is

$$\mathrm{E}_{\nu}[f(X)] = \int_{\mathcal{X}} f(x)\nu(x)dx.$$

By the transition kernel of the Markov process X, we mean

$$P(x, \mathcal{A}) = P\left(X_{n+1} \in \mathcal{A} | X_n = x\right), \quad \forall x \in \mathcal{X} \text{ and } \mathcal{A} \subset \mathcal{X}, \tag{1}$$

the probability of entering a set of states $\mathcal{A}$ at the next point in time, given that the process currently is in a state x. If there exist a distribution π, satisfying

$$\pi(\mathcal{A}) = \int_{\mathcal{X}} P(x, \mathcal{A})\pi(x)dx, \tag{2}$$

we say the π is a stationary distribution of π. The practical interpretation of π is that is represent the long term probability of being in a set of states $\mathcal{A}$, that is

$$\pi(\mathcal{A}) = \lim_{n\to\infty} P\left(X_n \in \mathcal{A}\right).$$

The product $\pi(x)dx$ in equation should be understood as the probability of being in a tiny set of size dx in $\mathcal{X}$. Showing the existence of a stationary distribution in general, can be quite hard, but it is important for us to recognize some circumstances for when a Markov process X has a unique stationary distribution. To establish this, we will now follow [10] very closely and start with the most important concept that must be fulfilled for a chain to have a stationary distribution, namely irreducibility.

Definition 2. *A Markov-process is φ-irreducible for a probability distribution φ on $\mathcal{X}$, if $\varphi(A) > 0$ for a set $\mathcal{A} \subset \mathcal{X}$ implies that*

$$P\left(\tau_{\mathcal{A}} < \infty | X_0 = x\right) > 0, \quad \forall x \in \mathcal{X},$$

where $\tau_{\mathcal{A}} = \inf_n\{X_n \in \mathcal{A}\}$. A Markov process is irreducible if it is φ-irreducible for some probability distribution φ. If a chain is φ-irreducible, then φ is called an irreducibility distribution for the process.

Defining irreducibility with respect to a probability distribution is just a technical detail when dealing with continuous state Markov processes. The essential meaning of a process being irreducible is the possibility of visiting any subset $\mathcal{A}$ in $\mathcal{X}$. A chain may have more than one irreducibility distribution, but it always has a maximum irreducibility distribution. A distribution ν_1 is absolutely continuous with respect to a distribution ν_2 if $\nu_2(\mathcal{A}) = 0 \Rightarrow \nu_1(\mathcal{A}) = 0$, for any $\mathcal{A} \subset \mathcal{X}$, and a maximal irreducibility distribution is such that all other irreducibility distributions are absolute continuous with respect to it. The next property we will look at is recurrence.

Definition 3. *An irreducible Markov process with maximal irreducibility distribution ψ is recurrent if for any set $\mathcal{A} \subset \mathcal{X}$ with $\psi(\mathcal{A}) > 0$ the conditions*

- $P\left(X_n \in \mathcal{A} \text{ infinitely often} | X_0 = x\right) > 0$ *for all* $x \in \mathcal{X}$

- $P\left(X_n \in \mathcal{A} \text{ infinitely often} | X_0 = x\right) = 1$ *for* $\psi - \text{almost all } x \in \mathcal{X}$

are both satisfied. An irreducible chain is positive recurrent if it has a stationary probability distribution. Otherwise it is null-recurrent.

While irreducibility was concerning the possibility of visiting any subset of states in the state space, recurrence is the property that this will happen over and over again. The definition introduces the concept of positive recurrence, as a property of irreducible chains with a stationary distribution. In the discrete state space case, it is usually the other way around, in the sense that if if all states are recurrent, the chain is both irreducible and positive recurrent. The theorem below is stated in [10] and conserns the uniqnueess of a stationary distribution.

Theorem 1. *Suppose the Markov process* $X = \{X_n, n \geq 0\}$ *is irreducible and that the distribution* π *satisfies equation* (2)*. Then the process is* π*-irreducible,* π *is a maximal irreducibility distribution,* π *is the unique stationary distribution of the chain, and the chain is positive recurrent.*

Note that theorem 1 does not tells us when process has a stationary distribution, but states that irreducibility is a requirement for uniqueness. Proving existence in the general case is beyond the scope of this thesis. However we will conclude in section 2.3, that if we can show that a process is regenerative, it will have a stationary distribution.

Before we end this section we define an important type of of Markov processes, called Lévy processes.

Definition 4. *A real valued process* $X = \{X_t, t \geq 0\}$ *with* $X_0 = 0$ *is a Lévy process if*

(i) X *has independent and stationary increments.*

(ii) X *is continuous in probability, that is, for a fixed* t, $\lim_{u \to t} P\left(|X_t - X_u| > \epsilon\right) = 0$, $\forall \epsilon > 0$.

We will use Lévy processes as degradation models in later sections. Lévy processes are contained in a broader class of process, called càdlàg processes. Càdlàg is a french abbreviation for "right continuous with left limits", and a càdlàg process have the property that

$$\lim_{u \to t^-} X_u < \infty \quad \text{and} \quad \lim_{u \to t^+} X_u = X_t.$$

When we use Lévy processes in degradation modeling, we often want them to be increasing. Increasing Lévy processes actually have their own name and are called subordinators.

2.2 Renewal processes

This section is mainly a summary of theory found in [22], where we only include theory relevant to us. By a counting process, we mean a stochastic process $\{N_t, t \geq 0\}$ counting the number of events by time t. A typical example of such a process may be the amount of customers having entered a store at some point during the day. Define S_i to be the time between the $(i-1)$st and ith event, known as inter-arrival intervals or sojourns. If the the sequence $\{S_i, i \geq 1\}$ of such inter-arrival times are independent and identically

distributed, we have by definition that N_t is a renewal process. Let F_S denote the cumulative distribution function of S_i, $i \geq 0$. In order to obtain important results on renewal process, it is appropriate to define U_n as the time of the nth renewal, with the convention that $U_0 = 0$. An illustration is provided in figure 1.

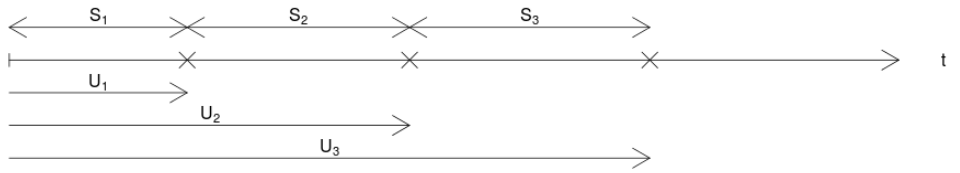

Figure 1: Illustration of sojourns and renewal times. Renewals are marked as crosses on the time axis.

Besides being an interesting quantity in itself, this is convenient since

$$N_t \geq n \Leftrightarrow U_n \leq t.$$

That is, we can express the distribution of N_t in terms of the distribution of U_n. Moreover since $U_n = \sum_{i=1}^{n} S_i$, the distribution of U_n is the n-fold convolution of F_S, denoted by F_S^{*n}. We then have

$$\begin{aligned} P(N_t = n) &= P(N_t \geq n) - P(N_t \geq n+1) \\ &= P(U_n \leq t) - P(U_{n+1} \leq t) \\ &= F_S^{n*}(t) - F_S^{(n+1)*}(t). \end{aligned}$$

Of course this distribution is typically hard to obtain. The function $m(t) = \mathrm{E}[N_t]$ is commonly referred to as the renewal function and it completely determines the renewal process. By conditioning on the time of the first renewal S_1, we have that

$$m(t) = \int_0^\infty \mathrm{E}[N_t | S = s] f_S(s) ds. \tag{3}$$

Obviously $\mathrm{E}[N_t | S = s] = 0$ if $s > t$. On the other hand if $s < t$, $\mathrm{E}[N_t | S = s] = 1 + E[N_{t-s}] = 1 + m(t-s)$, since one renewal will take place at time s and the process will then restart from here. Inserting this into (3), we get the so-called renewal equation

$$m(t) = 1 + \int_0^t m(t-s) f_S(s) ds. \tag{4}$$

In other words, by just knowing the distribution of the inter-arrival times $\{S_i, i \geq 1\}$, we may try to solve (4) to determine the renewal process completely. Unfortunately (4) is also typically hard to solve, but we may often recognize its structure. It is also possible obtain an explicit expression for $m(t)$ in terms of convolutions of F_S. Since N_t is positive and integer-valued we have

$$m(t) = \sum_{n=1}^{\infty} P(N_t \geq n) = \sum_{n=1}^{\infty} P(U_n \leq t) = \sum_{i=n}^{\infty} F^{n*}(t).$$

At time t we can notice that S_{N_t} is the time of the first renewal before t and S_{N_t+1} is the time of the next renewal (see figure 2), giving the double inequality

$$U_{N_t} \leq t < U_{N_t+1}. \tag{5}$$

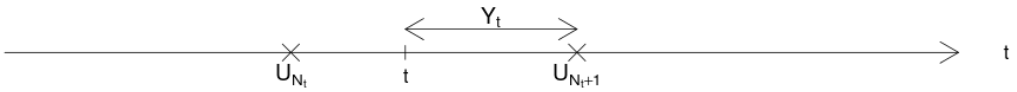

Figure 2: Illustration of S_{N_t} and S_{N_t+1}, the first renewals prior to and after time t. Y_t is then the time to the next renewal.

Dividing with N_t in (5), the left side becomes $\frac{U_{Nt}}{N_t} = \frac{1}{N_t}\sum_{i=1}^{N_t} S_i \to \mathrm{E}[S]$ as $t \to \infty$, since this is the mean of N_t sojourns. For the right side we similarly have that $\frac{U_{N_t+1}}{N_t} = \frac{U_{N_t+1}}{N_t+1} \cdot \frac{N_t+1}{N_t} \to \mathrm{E}[S] \cdot 1 = \mathrm{E}[S]$ as $t \to \infty$. Thus $\lim_{t\to\infty} \frac{t}{N_t} = \mathrm{E}[S]$, since it is squeezed in between the left and right side of (5) and we have the following

$$\lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{\mathrm{E}[S]}, \tag{6}$$

which is a classical result in renewal theory. It is also true for the renewal function $m(t) = \mathrm{E}[N(t)]$ that

$$\lim_{t\to\infty} \frac{m(t)}{t} = \frac{1}{\mathrm{E}[S]}. \tag{7}$$

The latter result is often known as the elementary renewal theorem. Although (6) and (7) looks similar, it is not the case that (7) follows from (6). We now show why (7) is true in the case of bounded sojourns, meaning that there is some finite real number $C > 0$ such that $P(S_i < C) = 1$ and $\mathrm{E}[S_i] < C$ for any i. To start out, we need a result known as Wald's equation, which says that if N is independent of a sequence of independent and identically distributed random variables $S_1, S_2, ...$, with $\mathrm{E}[N]$ and $\mathrm{E}[S_i] = E[S]$ being finite, then

$$\mathrm{E}\left[\sum_{i=1}^{N} S_i\right] = \mathrm{E}[N]\mathrm{E}[S]. \tag{8}$$

The proof of this can be found in many standard textbooks, like for example [22]. Now U_{N_t+1}, the time of the next renewal at time t, can be expressed in two ways. As before we may write $U_{N_t+1} = \sum_{i=1}^{N_t+1} S_i$ and apply Wald's equation (8) to obtain $\mathrm{E}[U_{N_t+1}] = \mathrm{E}[N_t + 1]\mathrm{E}[S] = (m(t) + 1)\mathrm{E}[S]$. On the other hand we may define Y_t as the time to the next renewal at time t (see figure 2) and write $U_{N_t+1} = t + Y_t$ and thus obtain $\mathrm{E}[U_{N_t+1}] = t + \mathrm{E}[Y_t]$. Equating the two expressions for $\mathrm{E}[U_{N_t+1}]$ and solving for $m(t)/t$ we get

$$\frac{m(t)}{t} = \frac{1}{\mathrm{E}[S]} + \frac{\mathrm{E}[Y_t]}{t\mathrm{E}[S]} - \frac{1}{t}. \tag{9}$$

Since $E[S] < C < \infty$ by assumption and $E[Y_t] < E[S]$ by construction, the middle term on the right hand side of (9) must be finite. Letting $t \to \infty$ equation (7) follows. It is also

possible that we earn some kind of reward during each sojourn. These rewards may also be negative, for example they may be interpreted as costs. Define R_i to be the reward earned between the $(i-1)$st and ith renewal, where it is possible that R_i will depend on the length of the sojourn S_i. Total accumulated rewards $R_t = \sum_{i=1}^{N_t} R_i$ are known as a Renewal-Reward process. Letting $E[R]$ be the mean reward, we then have that

$$\lim_{t\to\infty} \frac{R_t}{t} = \frac{\mathrm{E}\,[R]}{\mathrm{E}\,[S]} \tag{10}$$

and

$$\lim_{t\to\infty} \frac{\mathrm{E}\,[R_t]}{t} = \frac{\mathrm{E}\,[R]}{\mathrm{E}\,[S]}. \tag{11}$$

Under the specifications stated above, equation (11) is known as the Renewal-Reward theorem. Equation (10) is easily derived from the strong law of large numbers and equation (6). Since, as $t \to \infty$, $\frac{R_t}{N_t}$ is the mean reward and $\frac{N_t}{t} \to \frac{1}{\mathrm{E}\,[S]}$, we have $\frac{R_t}{t} = \frac{R_t}{N_t} \cdot \frac{N_t}{t} \to \mathrm{E}\,[R] \cdot \frac{1}{\mathrm{E}\,[S]}$. As with equation (6) and (7) it is also here the case that (11) does not follow from (10). To prove the Renewal-Reward theorem in the general case requires more theory than we want to include here, but we can show that it is correct in the case of independent and identically distributed rewards, being independent of N_t. In this case we can use Wald's identity (8) and have that $\frac{\mathrm{E}\,[R_t]}{t} = \frac{\mathrm{E}\left[\sum_{i=1}^{N_t} R_i\right]}{t} = \frac{\mathrm{E}\,[R]\mathrm{E}\,[N_t]}{t}$. As $t \to \infty$, we see from equation (7) that equation (11) follows.

Before leaving the topic of Renewal theory, we present a last type of processes called Markov Renewal processes.

Definition 5. *A stochastic process $(Y,T) = \{Y_n, T_n, n \geq 0\}$ with state space $\mathcal{Y} \times [0,\infty)$, with $T_0 = 0$ is called a Markov Renewal process if for any $\mathcal{A} \subset \mathcal{Y}$ and $t > 0$,*

$$\begin{aligned} &P\left(Y_{n+1} \in \mathcal{A}, T_{n+1} - T_n \leq t | Y_0, Y_1, ..., Y_n = y, T_0, T_1, ..., T_n\right) \\ &= P\left(Y_{n+1} \in \mathcal{A}, T_{n+1} - T_n \leq t | Y_n = y\right). \end{aligned} \tag{12}$$

Moreover the process (Y,T) is said to be a homogeneous Markov Renewal process if equation (12) is independent of n.

More loosely speaking, a Markov Renewal process is a two-dimensional process (Y,T), where further evolution given its history only depends on the present state of Y-process. The Y-process is thus a discrete time Markov process. Note that the next value of Y and the next increment of the T-process might be dependent. In later sections, we are going to see what a Markov Renewal process can look like.

2.3 Regenerative and semi-regenerative processes

With the background in the renewal theory from the previous section we will now discuss regenerative processes and semi-regenerative process. Loosely speaking a regenerative process is a stochastic process that with probability one will restart itself at some point in time, while a semi-regenerative process is at process that may restart itself at some point in time. We will see through its definitions that the regenerative process are as

closely linked to renewal processes, as the semi-renewal processes are linked to the Markov Renewal processes. In addition we will discuss the existence of stationary distributions for such processes. There are many equivalent ways to define a regenerative processes our definition is inspired by [3]. That is a regenerative process can be defined in terms of independent and identically distributed inter-arrival times at which the process starts over again.

Definition 6. *A Stochastic process $X = \{X_t, t \geq 0\}$ is said to be a regenerative process if there exists a renewal process with inter-arrival times $S_1, S_2, ...,$ with $\sup_n S_n = \infty$, such that the process $\{X_{S_n+t}, t \geq 0\}$ is independent of $S_1, ..., S_n$ and has the same distribution as the process $\{X_t, t \geq 0\}$.*

To be clear, a regenerative process is a process that evolves for some random amount of time, before it restarts itself. Possible values of X_0 and X_{S_n} is called the regeneration set, since the process will return to this set at the time of regeneration. After regeneration, the process will evolve according to the same distribution as it did from the start. Then it is quite intuitive to understand that the process is both irreducible and recurrent. For example, if we visit a set $\mathcal{A}$ during a sojourn, this must also be possible for some sojourn later, since the distribution of the process is the same in both periods. Moreover, since $\sup_n S_n = \infty$, we will visit $\mathcal{A}$ infinitely many times. Indeed, we have that according to [2], we have that if a process is regenerative, then it has a stationary distribution. Since the process is also irreducible, the stationary distribution is unique. It is often the case that we are not able to find an explicit solution to equation (2) and it therefore convenient to know if a unique solution exists. Another possible approach to check for uniqueness is to use analytical techniques from the theory of integral equations, given that we have an expression for the transition kernel P.

Following [7] we use the following definition for semi-regenerative processes.

Definition 7. *A càdlàg $\{X_t, t \geq 0\}$ is said to be a semi-regenerative process if there exist a Markov Renewal process (Y, T) with $\sup_n T_n = \infty$, such that the process $\{X_{T_n+t}, t \geq 0\}$ conditioning on $(T_0, ..., T_n, Y_0, ..., Y_n = y)$ have the same distribution as the process $\{X_t, t \geq 0\}$ given $Y_0 = y$. The process (Y, T) is called the embedded Markov Renewal process associated with $\{X_t, t \geq 0\}$.*

We see that this definition is in terms of càdlàg-processes, so remember that this is also valid Lévy processes as mentioned in section 2.1. What this definition really tells us, is that if we can show that a càdlàg process X has an embedded Markov Renewal process, the process X is a semi-regenerative process. The times $T_0, T_1, T_2, ...$ are called semi-regeneration times. Semi-regenerative processes is a generalization of regenerative processes. At the semi-regeneration times, the process either will restart itself or proceed as before until the next semi-regeneration point. [11] states the following important result for processes satisfying both definitions 6 and 7.

Theorem 2. *Let $X = \{X_t, t \geq 0\}$ be a Markov process with state space $\mathcal{X}$ which is both semi-regenerative with semi-regeneration times $\{T_n, n \geq 0\}$ and regenerative with regeneration times $\{S_n, n \geq 0\}$. Let $Y = \{Y_n, n \geq 0\} = \{X_{T_n}, n \geq 0\}$ be the embedded Markov process with stationary distribution π. Furthermore define $\Phi = \{\Phi_t, t \geq 0\}$, with $\Phi_0 = 0$ as a positive stochastic process with the properties that*

- $\Phi_t = \Psi_t(X_u, 0 \leq u \leq t)$
- $\Phi_t - \Phi_s = \Psi_{t-s}(X_u, s \leq u \leq t)$

for some function Ψ_t*. If for any* $t > 0$

$$\mathrm{E}\left[\Phi_t\right] < \infty \quad \textit{and} \quad \mathrm{E}\left[\Phi_{S_1}\right] < \infty,$$

then

$$\lim_{t\to\infty} \frac{\mathrm{E}\left[\Phi_t\right]}{t} = \frac{\mathrm{E}_{\,\pi}[\Phi_{T_1}]}{\mathrm{E}_{\,\pi}[T_1]}. \tag{13}$$

The process Φ is an additive function of the process X, like a reward process and theorem 2 is a generalization of equation (11). The notation $E_\pi[f(T_1)]$, means the expectation of some function f over a time T_1 given that that $Y_0 \sim \pi$. Equivalently we could have said the expectation of f over a semi-regenerative period. To simplify the notation we will therefore use the convention that

$$\mathrm{E}_{\,\pi}[f(T_1)] = \mathrm{E}_{\,\pi}[f(T)].$$

What theorem 2 is saying is that the expected long term rate at which the function Φ accumulates, is equal to the expected accumulation during a semi-regeneration period, divided by the expected length of the semi-regenration period, given that the state of the Y-process when the semi-regeneration start, is distributed according to it's stationary distribution. We will later see that Φ will represent accumulation of costs due to a maintenance program for a component. Theorem 2 will later be used extensively in computations of various estimates.

3 Maintenance theory and degradation models

3.1 Different types of maintenance

In this text we will not go into the details of maintenance theory, but give a short introduction to the concepts that is most important to us. We will assume that whenever a maintenance action takes place on an item, the item's state is reset so that item can be regarded to be "good as new" i.e. in perfect condition. In practice, such an action may be for example a replacement or a repair, but we will not look into these differences. However we want to distinguish between different reasons for a maintenance action to take place. [19] introduces the following distinctions between different types of maintenance:

- Preventive maintenance.
- Corrective maintenance.
- Failure-finding maintenance.

Failure-finding maintenance is a special class, that we will not consider in our model later on. It often consists of testing back-up and security solutions for a system, like for example a gas detector close to a pipeline system of flowing gas. The difference between preventive and corrective maintenance is of greater importance in this text and we will therefore take a closer look at these. Preventive maintenance (PM) is maintenance performed on an item while it is still functioning. Of course, this is done to prevent the event of an item failure. Many policies may be adopted to decide when a PM should be performed. [19] uses the the following classification:

- Age-based maintenance.
- Clock-based maintenance.
- Condition-based maintenance.
- Opportunity maintenance.

Age-based and clock-based maintenace are somehow similar. Using an age-based policy for PM, the maintenance is performed when the item reaches a certain age, like for example 3 years, while using clock-based policy, the PM is performed a pre-specified calendar times like for example every third week. Under the condition-based policy, PM is performed based on measurements of the items condition. We will comment on such measurements later in this section. Opportunity maintenance is only relevant for systems consisting of more than one item, and basically means that a PM is performed based on an opportunity to do so, typically because maintenance is performed on a neighboring component. An example may be a valve and a filter inside a pipeline. If the system is shut down because the valve has to maintained, one might want to maintain the filter as well. We will mainly be concerned with condition-based maintenance in applications later.

Corrective maintenance (CM) is maintenance performed on an item after it has failed. Maintenance actions are also associated with costs, that we will discuss in details in section 4. For now, we note that if a preventive maintenance regime should make sense, there must be a possibility to reduce total costs due to maintenance under this regime, compared to a

regime where we let items run until failure. In connection with condition-based preventive maintenance, we should remark on how we could keep track on an item's condition. First of all this requires at least one measurable variable being correlated with or representing the item's state. [19] discusses different possibilities here, including physical variables (like temperature and pressure) and performance variables (for example quality of produced items) amongst others. We will assume that the system's state can be summarized in one continuous random variable and that the time evolution of this can be a described by a degradation process (see next section). The next question is how to obtain information on an item's state. Typically this can either be done by inspection or by monitoring. Monitoring means that we have a device providing real time information on the item's state, such as a camera or a temperature sensor. This can be done either continuously or sequentially. Inspection, which will be our approach, means that we gather information on the item's state at certain times by examination of the condition variable. Inspections is typically done at pre-specified times. These time points may be equidistant, for example 8 o'clock every morning, or they may be determined based on the item's condition at last inspection.

3.2 Degradation processes

Degradation processes, also sometimes referred to as deterioration processes, are stochastic processes modeling the state of an item subject to random degradation. For a more complete introduction and background on this topic, see for example [24]. If we let $X = \{X_t, t \geq 0\}$ be a degradation process, we will use the convention that $X_t = 0$ means that the item is in perfect condition. Over time the item will continuously degraded until it reaches a level $L > 0$, so that when $X_t \geq L$, the item is considered to be in a failed state. In practice this means that the use of degradation processes, requires a that we are able to measure the item's state, and hence its degradation, which can be a challenging task. However this problem will not be dealt with in this text and we will assume that this is always possible. For now, we are considering a situation where no kinds of maintenance is performed. We then define the lifetime of the item as

$$T_L = \inf_t \{X_t \geq L\}. \tag{14}$$

By a lifetime distribution we mean the distribution of T_L. A classical approach in reliability engineering is to consider the lifetime distribution of components that are assumed to be in one of two states: the functioning state or the failed state. Making use of degradation processes, we are able to study items that might be in a whole range of different states. A graphical interpretation of the framework presented so far can be seen in figure 3, where we display a sample path of a typical degradation process and have labeled the failure level L and the lifetime T_L.

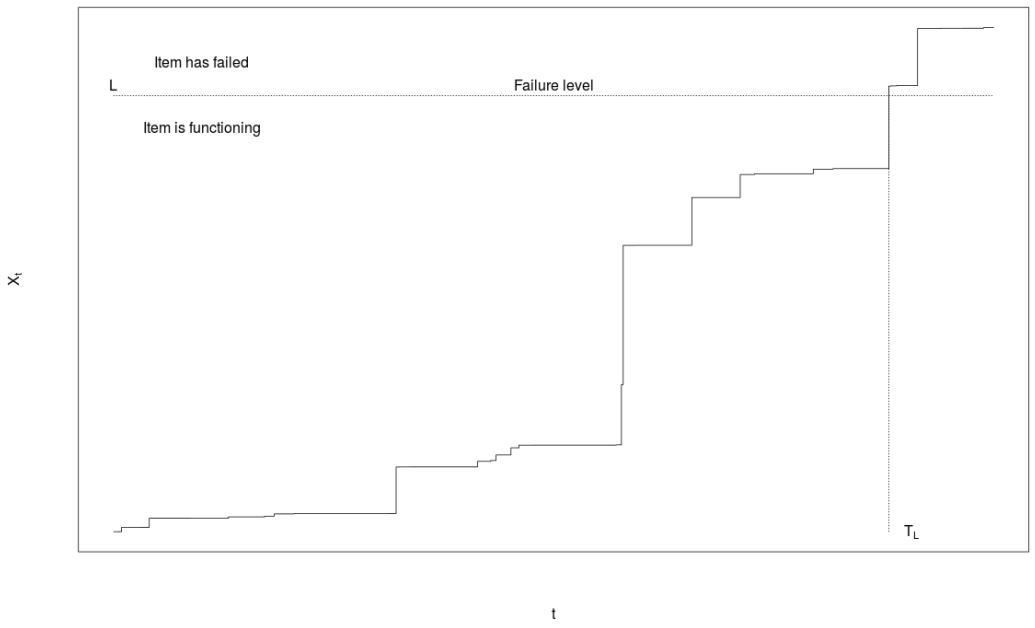

Figure 3: Sample path of a typical degradation process.

For reasons that will become clear later, it should be mentioned that reaching level L, does not necessarily mean that we will detect the failure when it happens. Therefore the time spent above L will be referred to as the downtime of the system. Depending on whether or not the process is below L, we see that we can still apply the classical two-state approach, although for some processes it can be hard to obtain the lifetime distribution. If the degradation process X is an increasing process, we can attempt to find the distribution of T_L from

$$P(T_L < t) = P(X_t > L). \tag{15}$$

Among the desired properties a degradation process should have, it is often preferred that the process is increasing in the absence of maintenance. The Wiener process is an example of process that does not have this property. Although it is sometimes used to model degradation by including a positive drift in the model, it's negative increments may be diffucult to interpret physically. Another property often desidered by reliability engineers is that the process can be expressed as a positive Compound Poisson process [26]. That is a process $\sum_{i=1}^{N_t} D_i$, where N_t is a Poisson process and the D_i's are positive identically distributed random variables, often referred to as jumps. If a degradation process can be expressed as a positive Compound Poisson process, the degradation at time t can be interpreted as a series of random shocks causing the degradation. We will now present two particular degradation processes that has these properties. As we will see these processes are specified by certain parameters that we will assume to be constant. The models can be extended by letting these parameters being functions of time, without too much complication. Another approach used in [26] is the let the parameters be random variables.

3.2.1 The Gamma process

The Gamma process is a very common model for degradation and has been found to be realistic in some applications [19]. [18] is an excellent summary of both properties and applications of the Gamma process. Before defining the process, we need to introduce the Gamma distribution (Ga). If a random variable $X \sim \text{Ga}(\alpha, \beta)$, then X has a probability density function

$$f_X(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \alpha > 0, \beta > 0, \tag{16}$$

where $\Gamma(a) = \int_0^\infty u^{a-1} e^{-u} du$ is known as the gamma function. α is called the shape parameter and β is called the scale parameter. One might also specify the model in terms of α and $\tilde{\beta} = \beta^{-1}$. The cumulative distribution function is

$$F_X(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t\beta} dt = \frac{1}{\Gamma(\alpha)} \int_0^{\beta x} u^{\alpha-1} e^{-u} du = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}, \tag{17}$$

where $\gamma(a, b) = \int_0^b u^{a-1} e^u du$ is known as the lower incomplete gamma function. From (16) we can find that

$$\text{E}[X] = \frac{\alpha}{\beta} \quad \text{and} \quad \text{Var}(X) = \frac{\alpha}{\beta^2}. \tag{18}$$

The Gamma process is then defined as follows

Definition 8. *A stochastic process $X = \{X_t, t \geq 0\}$ is said to be a Gamma process with shape parameter α and scale parameter β if*

(i) $X_0 = 0$.

(ii) *X has independent and stationary increments.*

(iii) *For every $t > s$, $X_t - X_s \sim Ga(\alpha(t-s), \beta)$.*

Since the density of the Gamma distribution has positive support, the Gamma process is a subordinator. In figure 4 we display the graph of different Gamma density functions with a sample path of its associated Gamma process below.

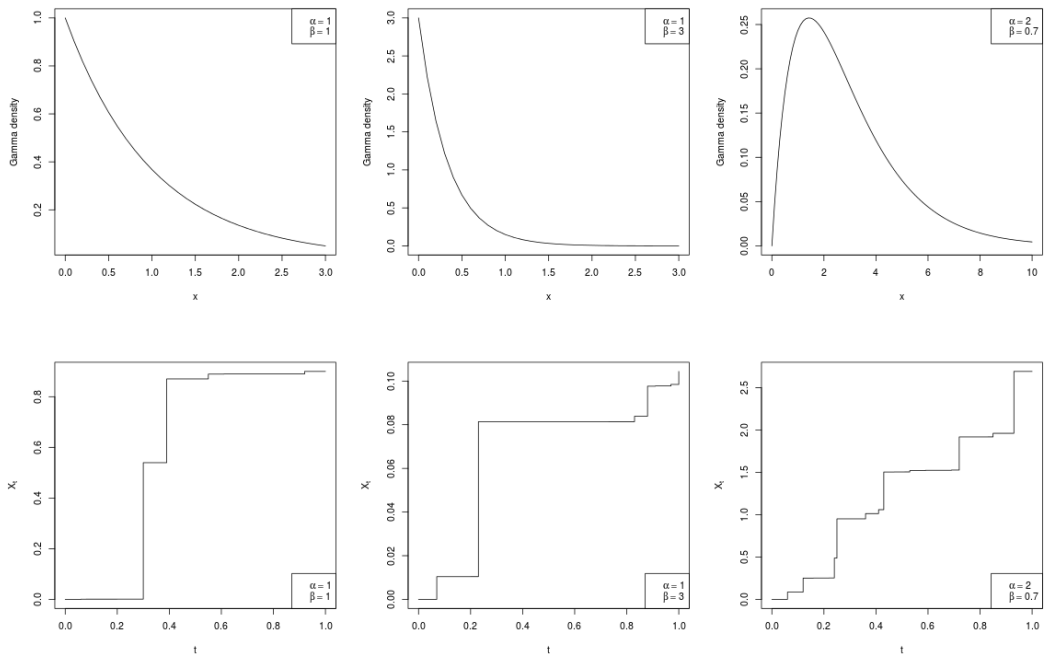

Figure 4: Gamma densities and associated Gamma processes for different choice of parameters.

[12] argues that the Gamma process can be thought of as a compound Poisson process with Gamma distributed jumps and an arrival rate which is going to infinity. For the distribution of the lifetime T_L, defined by equation (14), we get from (17) and (15), that

$$P(T_L < t) = 1 - \frac{\gamma(\alpha t, \beta L)}{\Gamma(\alpha t)}. \tag{19}$$

As pointed out by [18], differentiation of (19) does not give a closed form expression for the density of T_L, but it can be expressed in terms of socalled digamma functions. This means that in order to do computations on this distribution, a numerical algorithm is required.

3.2.2 The Inverse Gaussian process

Another degradation model that recently have become more popular is the Inverse Gaussian process [25],[26]. Again we have to introduce the Inverse Gaussian distribution (IG) before defining this process. [6] and [23] are both books devoted to this distribution. If a random variable $X \sim \mathrm{IG}(\mu, \eta)$, then X has the probability density function

$$f_X(x; \mu, \eta) = \sqrt{\frac{\eta}{2\pi}} x^{-3/2} \exp\left(-\frac{\eta(x-\mu)^2}{2\mu^2 x}\right), \quad x > 0, \mu > 0, \eta > 0. \tag{20}$$

μ is called the mean parameter and η is called the scale parameter. Other parametrizations are also possible (see [23]). A nice property with the IG-distribution is the its cumulative distribution function can be written in terms of the cumulative distribution function

$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-t^2/2} dt$, of the standard normal distribution. This is proved in [6] and the expression is

$$F_X(x; \mu, \eta) = \Phi\left[\sqrt{\frac{\eta}{x}}\left(\frac{x}{\mu} - 1\right)\right] + e^{2\frac{\eta}{\mu}}\Phi\left[-\sqrt{\frac{\eta}{x}}\left(1 + \frac{x}{\mu}\right)\right]. \tag{21}$$

From (20) we can find that

$$\mathrm{E}[X] = \mu \quad \text{and} \quad \mathrm{Var}(X) = \frac{\mu^3}{\eta}. \tag{22}$$

The IG-process can now be defined.

Definition 9. *A stochastic process $X = \{X_t, t \geq 0\}$ is said to be an Inverse Gaussian process (IG-process) with mean increment μ and scale paramteter η if*

(i) $X_0 = 0$.

(ii) X_t *has independent and stationary increments.*

(iii) For every $t > s$, $Z_t - Z_s \sim IG((t-s)\mu, (t-s)^2\eta)$.

As we can see from this definition and equation (20), the IG-process is also a subordinator. Figure 5 shows graphs of different IG-densities with a sample path of its associated IG-process below.

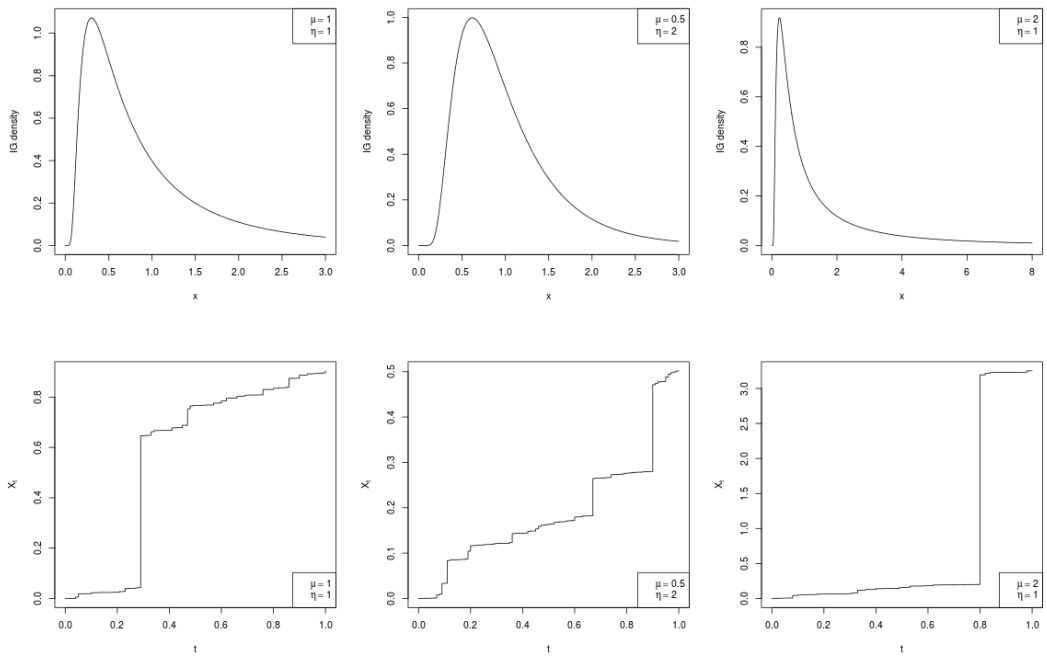

Figure 5: IG densities and associated IG-processes for different choice of parameters.

[26] proves that the IG-process can be expressed as a compound Poisson process. Moreover, using (21), we get that the lifetime distribution for an item under this degradation process is

$$P(T_L < t) = \Phi\left[\sqrt{\frac{\eta}{L}}\left(t - \frac{L}{\mu}\right)\right] - e^{2\frac{t\eta}{\mu}}\Phi\left[-\sqrt{\frac{\eta}{L}}\left(t + \frac{L}{\mu}\right)\right] \tag{23}$$

Since the right hand side of (23) is expressed in terms of the standard normal cumulative distribution function, we are able to obtain an analytic expression for the density of T_L. [26] have two specific remarks to equation (23). Firstly, they show that the distribution of T_L can be approximated by a Birnbaum-Saunders distribution, which is one of the classical lifetime distributions in reliability engineering (see [19]). This partly builds up under the choice of using the IG-process as a degradation model. The second remark, is that since we are able to obtain an analytical expression for the density of T_L, the IG-process has an advantage compared to the Gamma process when it comes to estimation. This is because the lack of a closed form expression for the lifetime distribution in the Gamma case, makes exact statistical inference analytically intractable.

4 A complete model for degradation and maintenance

We now present a degradation and maintenance model given in [11] and [9]. At first we will follow the these papers very closely, before we will discuss possible modifications. The system in this model consists of just one item, where the item's state at time t can be summarized in one random variable X_t.

4.1 The original model

Without any forms of maintenance X_t will be a strictly increasing process, namely the Gamma process presented in section 3.2.1. It is assumed that the system is in a failed state when $X_t \geq L$, for some fixed level L. When the system is in perfect condition (good as new), we have $X_t = 0$. The system is not monitored, so the only way we can get information about its state X_t is by inspection. These inspections are assumed to be perfect, in the sense that we observe the system's state without error in a negliable amount of time. Failures are also detected only through inspections. Having the observed the system's state, one of the following decisions must be made:

- Perform corrective maintenance.
- Perform preventive maintenance.
- Leave the item as it is.

In addition a date for the next inspection must be set. The times at which the system is inspected will be denoted $T_1, T_2, ...$ The corrective and preventive maintenance could be both repairs or replacement. Nevertheless these actions are also assumed to be perfect, that is, they take no time and resets the item to the "good as new" state. Letting $X_{T_n^-}$ be the item's state just before inspection number n, the maintenance decisions are made using the following rules:

- If $X_{T_n^-} \geq L$, corrective maintenance is performed.
- If $M \leq X_{T_n^-} < L$, for some level $M < L$, preventive maintenance is performed.
- Otherwise, the item is left as it is.

We use the convetion that at the time of the nth inspection, the item's state is

$$X_{T_n} = \begin{cases} 0 & \text{if } X_{T_n^-} > L \\ 0 & \text{if } M \leq X_{T_n^-} < L \\ X_{T_n^-} & \text{otherwise.} \end{cases} \tag{24}$$

To determine the time for the next inspection we use the function

$$m(x) = 1 + \max\left[A\left(1 - \frac{x}{B}\right), 0\right], \quad A > 0, B > 0, \tag{25}$$

such that

$$T_{n+1} = T_n + m\left(X_{T_n}\right). \tag{26}$$

We will comment on the coefficients A and B later. Figure 6 illustrates how equation (25) and (26) are used to determine the next inspection time. Note that the next inspection time is dependent only on the system state at the current inspection.

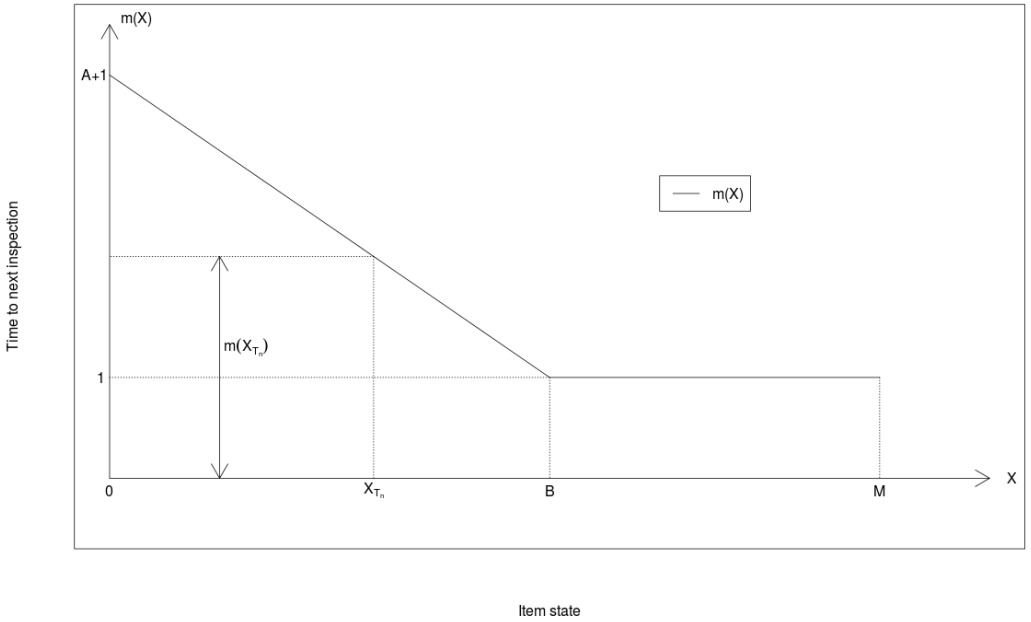

Figure 6: Graph of function m given by (25) and the determination of the next inspection time given by (26).

The choice of the function m can actually be quite general. According to the model, any positive and decreasing function $m : [0, M) \to [m_{\min}, m_{\max}]$, with $m_{\max} = m(0) > m_{\min} = m(M) > 0$ can be used in equation (26). This ensures that the sequence $\{T_n, n \geq 0\}$ will be increasing and that the time to next inspection will become shorter the more we approach a maintenance action. We define also the sequence of maintenance times, $S_1, S_2, ...$, where the replacements can be both preventive or corrective. A illustration of the proposed framework can be seen in figure 7.

Let us now take the following costs into account:

- C_i, the cost of one inspection.
- C_p, the cost of a preventive maintenance action.
- C_c, the cost of a corrective maintenance action.
- C_d, the cost of downtime pr. time unit.

As mentioned in section 3.1 there must be a possible gain by introducing a preventive maintenance regime. To be absolute certain of this, we will assume $C_p < C_c$. Let $N_t^{(i)}$ be the numbers of inspections performed, $N_t^{(p)}$ be the numbers of preventive maintenance

actions taken, $N_t^{(c)}$ be the number of corrective maintenance actions taken and d_t be the total accumulated downtime in the interval $[0, t]$. Define the global cost process as

$$C_t = C_i \cdot N_t^{(i)} + C_p \cdot N_t^{(p)} + C_c \cdot N_t^{(c)} + C_d \cdot d_t \tag{27}$$

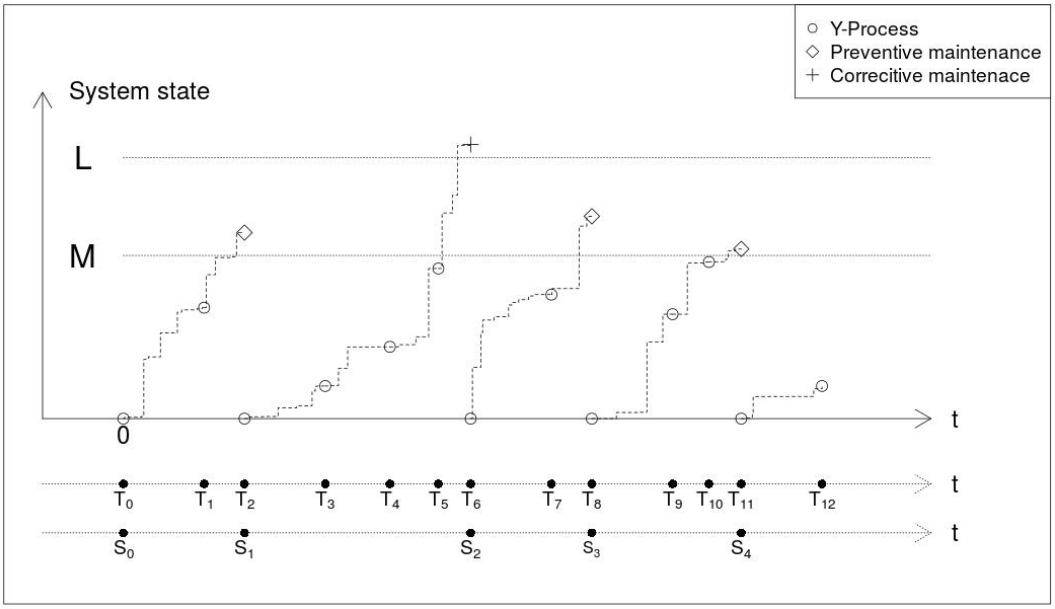

Figure 7: Illustration of the of the model. Sample paths of the unobserved degradation process between inspections is represented by the dashed lines.

Now our benchmark of how well a certain maintenance strategy is functioning will be the expected global cost rate at an infinite horizon, denoted by EC_∞ and defined via (27) as

$$\begin{aligned} EC_\infty &= \lim_{t\to\infty} \frac{\mathrm{E}\left[C_t\right]}{t} \\ &= \lim_{t\to\infty} \left(C_i \cdot \frac{\mathrm{E}\left[N_t^{(i)}\right]}{t} + C_p \cdot \frac{\mathrm{E}\left[N_t^{(p)}\right]}{t} + C_c \cdot \frac{\mathrm{E}\left[N_t^{(c)}\right]}{t} + C_d \cdot \frac{\mathrm{E}\left[d_t\right]}{t} \right). \end{aligned} \tag{28}$$

We want of course EC_∞ to be as small as possible. Our decision variables are M, A and B. That is, we can choose when we want to start carrying out preventive maintenance actions in the variable M and we can choose how often we want to do inspections in the variables A and B. [11] argues that the limits occurring in (28) is hard to determine, since they are functions of the degradation process X before time t. However, we are now going to explore properties of this model and define convenient processes that can help us compute (28) efficiently. To start out, notice that $S_1, S_2 - S_1, ...$ are independent and identically distributed. This follows from the fact that the degradation process X is a Markov process (which follows from the definition of the Gamma process). We can therefore identify the global cost process C_t as a renewal-reward process (see section 2.2) and use equation (11) to establish

$$EC_\infty = \frac{\mathrm{E}\left[C_S\right]}{\mathrm{E}\left[S\right]}, \tag{29}$$

where $\mathrm{E}[f(S)]$ denotes the expectation of some function f over one sojourn. In other words, the long term expected cost rate is equal to the expected cost during a renewal cycle, divided by the expected length of the problem. Now equation 29 is much easier to compute than equation (28), but with the proposed model we can actually do better. Define the process $Y = \{Y_n, n \geq 0\} = \{X_{T_n}, n \geq 0\}$, the observations of X at the inspection dates. Clearly Y is an embedded Markov process, having the state space $[0, M)$. Two very important properties of the degradation process X will now be stated:

(i) X_t is regenerative, with regeneration times $\{S_n, n \geq 0\}$.

(ii) X_t is semi-regenerative with semi-regeneration times $\{T_n, n \geq 0\}$.

(i) follows more or less from the construction of this model. After each maintenance action, the item is restored to a "good as new"-state, or in mathematical terms, at the times $\{S_n, n \geq 0\}$, we have $X_{S_n} = 0$, and by the Markov property of X, X_{S_n+t} must have the same distribution as X_t, meaning that statement (i) is true (see definition 1, section 2.3). To see that statement (ii) is true, we need to consider the process (Y, T) and see why this is a Markov Renewal process, as defined in section 2.2. In other words, we need equation (12) to be true for the process (Y, T) defined in this model, stating that the next observation $(Y_{n+1}, T_{n+1} - T_n)$ is only dependent on Y_n. From equation (25) we see that this holds for the time increment, since it can be rearranged to

$$T_{n+1} - T_n = m(X_{T_n}) = m(Y_n).$$

Moreover $Y_{n+1}|Y_n \sim Ga(\alpha \cdot m(Y_n), \beta)$ is also dependent on Y_n only. Therefore we have that (Y, T) is a Markov Renewal process and it then follows from definition 8, section 2.3 that X_t is a Semi-regenerative process with semi-regeneration times $\{T_n, n \geq 0\}$ as stated in (ii). The fact that X is both regenerative and semi-regenerative is important because of its consequences. First of all, the regenerative property alone implies that X has a stationary distribution, and since Y is embedded in X, so does Y. Let π denote the stationary distribution of Y. Since $Y_{S_n} = 0$, that is the process returns to the "good as new"-state after maintenance, $\{0\}$ is called a regeneration set for Y and we can assume the density of π is on the form

$$\pi(y) = a \cdot \delta(y) + (1 - a) \cdot b(y), \tag{30}$$

where $\delta(y)$ is the Dirac delta, $b(y)$ is a probability density function and $0 < a < 1$. (30) basically says that a is the long term probability for the process being in the state 0 and $1 - a$ is the long term probability of observing the chain in some other state in $(0, M)$, where the distribution of these are given by $b(y)$. Since we have that X is regenerative, semi-regenerative and the embedded Markov process Y has a stationary distribution, it is tempting to invoke theorem 2, section 2.3 to improve on equation (29). We then let $\Phi_t = C_t$ and check if all the assumptions stated for Φ_t also hold for C_t. Clearly C_t is an additive function of the process X, so the two first properties are fulfilled. To see if $\mathrm{E}[C_t] < \infty$, we must check this for all the processes $N_t^{(i)}$, $N_t^{(p)}$, $N_t^{(c)}$ and d_t in (27). Since the number of inspections is larger or equal to the number of maintenance actions we have that $N_t^{(p)} \leq N_t^{(i)}$ and $N_t^{(c)} \leq N_t^{(i)}$. To see that $N_t^{(i)}$ has a bound for every finite t, recall

that in the general case, the function m that determines the time to the next inspection has a lower bound $m_{\min}$. From this we deduce that for any integer $k > 0$, we may write

$$t = T_1 + (T_2 - T_1) + ... + (T_k - T_{k-1}) + (t - T_k) \geq m_{\min} \cdot (N_t^{(i)} + 1) > m_{\min} \cdot N_t^{(i)}.$$

Therefore $N_t^{(i)} < \frac{t}{m_{\min}}$ and we have $\mathrm{E}\,[N_t^{(i)}] < \infty$, $\mathrm{E}\,[N_t^{(p)}] < \infty$ and $\mathrm{E}\,[N_t^{(c)}] < \infty$. Moreover $d_t \leq t$, with equality if and only if the item is non-functioning for all t. Therefore $E[C_t] < \infty$, and we have showed that this condition is satisfied for the global cost function. Let S be the time of the first maintenance action. It remains to show that $E[C_S] < \infty$. In this case we can tell that $\mathrm{E}\,[N_S^{(p)}] \leq 1$ and $E[N_S^{(c)}] \leq 1$, since we at maximal can perform one maintenance action in each regeneration period. Define $\tau = \inf_{n\geq 1}\{Y_n = 0\}$. We then have that $N_S^{(i)} = \tau$ and that $d_S \leq S \leq m(0) \cdot \tau$. Now S can be viewed as a passage time of the Gamma process, and since the Gamma process is strictly increasing it will sooner or later reach a point where a maintenance action is taken. Therefore $\mathrm{E}\,[S] < \infty$. Moreover, we have that $S \geq m_{\min} \cdot \tau$, so that $\mathrm{E}\,[\tau] \leq \frac{\mathrm{E}\,[S]}{m_{\min}} < \infty$. Thereby $\mathrm{E}\,[N_S^{(i)}] = \mathrm{E}\,[\tau] < \infty$ and $\mathrm{E}\,[d_S] \leq \mathrm{E}\,[S] < \infty$. We conclude that $\mathrm{E}\,[C_S] < \infty$ and that we therefore can apply theorem 2, section 2.3 to obtain

$$EC_\infty = \frac{\mathrm{E}_\pi[C_T]}{\mathrm{E}_\pi[T]} = C_i \cdot \frac{\mathrm{E}_\pi[N_T^{(i)}]}{\mathrm{E}_\pi[T]} + C_p \cdot \frac{\mathrm{E}_\pi[N_T^{(p)}]}{\mathrm{E}_\pi[T]} + C_c \cdot \frac{\mathrm{E}_\pi[N_T^{(c)}]}{\mathrm{E}_\pi[T]} + C_d \cdot \frac{\mathrm{E}_\pi[d_T]}{\mathrm{E}_\pi[T]} \qquad (31)$$

Now, the best thing about (31) is that we actually can derive intuitive expressions for many of the quantities involved. And even more important, we can establish natural estimators for all of the quantities from the output of a Monte Carlo simulation of the (Y, T) process. Moreover, in such simulations, we can express all these quantities in terms of the first inspection T_1, given that we start the process in a value sampled from the stationary distribution π. We have for example that

$$\mathrm{E}_\pi[N_T^{(i)}] = 1,$$

since we only do one inspection between $(T_{k-1}, T_k]$, for any $k = 1, 2, 3...$, and does not need to be estimated. We also have that

$$\mathrm{E}_\pi[N_T^{(p)}] = P_\pi(M \leq X_{T_1^-} < L) \quad \text{and} \quad \mathrm{E}_\pi[N_T^{(c)}] = P_\pi(X_{T_1^-} \geq L),$$

where $P_\pi(\cdot)$ means probabilities computed in terms of the stationary distribution π, requiring of course that $X_{T_0} \sim \pi$. Similar, but more complicated, expressions can also be derived for $\mathrm{E}_\pi[T]$ and $\mathrm{E}_\pi[d_T]$, but we will not use them. Instead, the main idea now is to establish quantities that can estimate all the ingredients in (31), during a Monte Carlo simulation of (Y, T). In section 5, we will explain such a simulation algorithm in detail. For now, we assume that such an algorithm for (Y, T) is available. Let A be any event, we then denote by $\#(A)$, then numbers of times A have occurred. The estimators we will use for computing (31) is then

$$\widehat{E_\pi[N_T^{(p)}]} = \frac{\#(M \leq X_{T_n^-} < L)}{\#(\text{inspections})}. \qquad (32)$$

$$\widehat{E_\pi[N_T^{(c)}]} = \frac{\#(X_{T_n^-} > L)}{\#(\text{inspections})}. \qquad (33)$$

$$\widehat{\mathrm{E}_\pi[d_T]} = \frac{\text{Total accumlated downtime}}{\#(\text{inspections})} \tag{34}$$

$$\widehat{\mathrm{E}_\pi[T]} = \frac{\sum_{n=1}^{\#(\text{inspections})} (T_n - T_{n-1})}{\#(\text{inspections})}, \tag{35}$$

with $T_0 = 0$. These will all converge, due to the law of large numbers. To make sure that the simulations are started with an initial value $Y_0 \sim \pi$, a burn-in period is used.

4.2 Possible extensions

An advantage by taking the Monte Carlo approach for simulation is that we can easily add extra elements into the model, without having to derive new equations. Many such extensions can be made, without violating the assumptions of theorem 2 from section 2.3, so that the estimates (32)-(35) are still valid. We now present two such extended models.

4.2.1 Changing the degradation model

Adapting a new model for the degradation of the item is one of the easiest extensions we can make. This is also a very nice opportunity to have, since we may find other models appropriate than the Gamma process. The only thing we have to make sure of is that the process is a cádlág and that it is regenerative under the inspection/maintenance-program we have introduced. We have chosen to use the Inverse Gaussian process as an alternative degradation process in our numerical example, that will be presented later.

4.2.2 Non-perfect inspections

Another possible approach is to introduce non-perfect inspections. In what we have presented so far, we have been able to obtain the information about the system's state at inspection without any kind of error. Different approaches could be used to model non-perfect inspections, and we use a model inspired by [14]. Here one assumes that a signal is emitted from the item when the process X reaches the level M. The signal is detected with probability q and a preventive maintenance action is taken. If the signal is not detected, the item is left as it is until the next inspection.

5 Simulation methods

In this section we will construct a Monte Carlo-based algorithm for simulating the model we presented in section 4. To be more precise we will have to simulate the (Y, T)-process, as this is the process that will be observed in practice. A problem that then occurs is how to estimate the accumulated downtime d_t of the system, since the actual passage time of the level L for the degradation process will not be observed. To be clear, we may for example make two observation $Y_{n-1} < L$ and $Y_n > L$ at the times T_{n-1} and T_n respectively. We then know that there is a passage time $T_L \in (T_{n-1}, T_n)$ and the there is a downtime $T_n - T_L$ associated with this failure. To solve this problem, we will construct an algorithm for drawing from the distribution of these passage times, conditioning on the observations of the Y-process prior to, and after a failure has occurred. We will see that we can solve this by a bisection algorithm that is heavily based on a concept called bridge sampling. Therefore we will start out by introducing bridge samplers and then develop such samplers for both the Gamma process and the IG-process. After having established the bridge samplers we will put the pieces together to the algorithm for drawing passage times. Towards the end of the section, we will put the passage time algorithm into the big picture, the entire model. A short discussion on the convergence of Monte Carlo algorithm is also included.

5.1 Bridge Sampling

Let $X = \{X_t; t \geq 0\}$ be a Lévy process and consider three time points $t_\ell < t < t_u$. Bridge sampling is a method to simulate realizations of X_t, that is the bridge point, given that we know the value of the process at some earlier time t_ℓ and at some time later t_u. The way this is done is to first obtain the distribution of $X_t - X_{t_\ell}$ conditioning on $X_{t_u} - X_{t_\ell}$, known as the bridge distribution. Having the bridge distribution at hand, we may draw realizations of the bridge point directly if possible or, as we shall see in in examples, obtain X_t exploiting the bridge distribution by clever use of transformations. In what follows define the random variables for the increments of the process

$$W = X_t - X_{t_\ell} \quad , \quad Y = X_{t_u} - X_t \quad \text{and} \quad Z = X_{t_u} - X_{t_\ell} \tag{36}$$

and for the time increments let

$$\tau_w = t - t_\ell \quad , \quad \tau_y = t_u - t \quad \text{and} \quad \tau_z = t_u - t_\ell.$$

Obviously $Z = W + Y$ and $\tau_z = \tau_w + \tau_y$. Figure 8 provides an illustration of the entire setting.

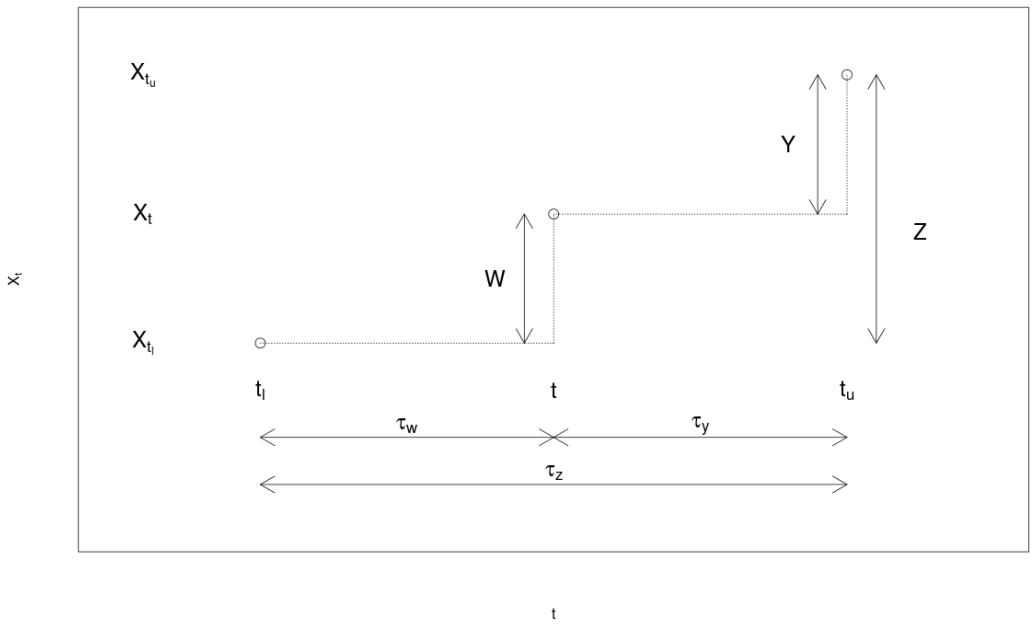

Figure 8: Bridge sampling and some of the quantities involved

Conditioning on Z, the distribution of W is

$$f_{W|Z}(w) = \frac{f_{W\cap Z}(w,z)}{f_Z(z)} = \frac{f_{Z|W}(z)f_W(w)}{f_Z(z)} = \frac{f_Y(z-w)f_W(w)}{f_Z(z)} \tag{37}$$

Now given a Lévy process X, equation (37) is an expression for the bridge distribution using only the known distributions of the increments. Therefore the first step in order to obtain a bridge sampling algorithm for a process is to compute the right hand side of this equation. We proceed by deriving such algorithms for both the Gamma process and the Inverse Gaussian process. It turns that bridge simulations of both these processes are quite elegant and that the construction the Gamma bridge is much easier than the Inverse Gaussian bridge, which makes it a very good introductory example.

5.1.1 Gamma bridge sampling

The source of this derivation is [20]. Letting X be the Gamma process, meaning $X_t - X_s \sim Ga(\alpha(t-s), \beta)$, equation (37) becomes

$$f_{W|Z}(w) = \frac{\Gamma(\alpha\tau_w + \alpha\tau_y)}{\Gamma(\alpha\tau_w)\Gamma(\alpha\tau_y)} \cdot z^{-1} \left(\frac{w}{z}\right)^{\alpha\tau_w - 1} \left(1 - \frac{w}{z}\right)^{\alpha\tau_y - 1} \tag{38}$$

Consider now (for Z given) the transformation $P = \frac{W}{Z}$ have the distribution

$$f_P(p) = f_{W|Z}(zp) \cdot z = \frac{\Gamma(\alpha\tau_w + \alpha\tau_y)}{\Gamma(\alpha\tau_w)\Gamma(\alpha\tau_y)} p^{\alpha\tau_w - 1}(1-p)^{\alpha\tau_y - 1}.$$

That is P has a Beta-distribution $P \sim Be(\alpha\tau_w, \alpha\tau_y)$. More background on the Beta distribution can be found in many standard textbooks, for example [4]. Finally we note

that from the definitions (36) of W and Z that we can write $P = \frac{X_t - X_{t_\ell}}{X_{t_u} - X_{t_\ell}}$, giving

$$X_t = X_{t_\ell} + P \cdot (X_{t_u} - X_{t_\ell}) \tag{39}$$

Assuming that we are able to sample from the Beta-distribution, a summary of the bridge sampling algorithm is as follows:

Data: t,t_ℓ,t_u,X_{t_ℓ}, X_{t_u} and α
Result: $X_t | X_{t_\ell}, X_{t_u}$
Draw $P \sim Be\left(\alpha(t - t_\ell), \alpha(t_u - t)\right)$;
$X_t \leftarrow X_{t_\ell} + P \cdot (X_{t_u} - X_{t_\ell})$;
return X_t

Algorithm 1: Gamma bridge point

By successively running this algorithm an entire sample path can be simulated. For example if we generate X_t, from X_{t_ℓ} and X_{t_u}, we can generate two new bridge points using X_{t_ℓ} and X_t for the first, and X_t and X_{t_u} for the second, and so on. As pointed out by [20] there is a very nice structure in this algorithm that can be seen from equation (39). Since P has a Beta distribution we have that $P \in [0, 1]$. The bridge point X_t is therefore a sum of the lower value X_{t_ℓ} and a proportion of the entire increment $X_{t_\ell} - X_{t_u}$ of the process from t_ℓ to t_u, which is somewhat intuitive taking prior knowledge about the Gamma process into account. Below, in figure 9, we display some simulated sample paths that can be compared with the sample paths of figure 4 in section 3.2.1, since the parameters are the same. All these simulations are done with $[t_\ell, t_u] = [0, 1]$ with $X_0 = 0$ and $X_1 \sim Ga(\alpha, \beta)$.

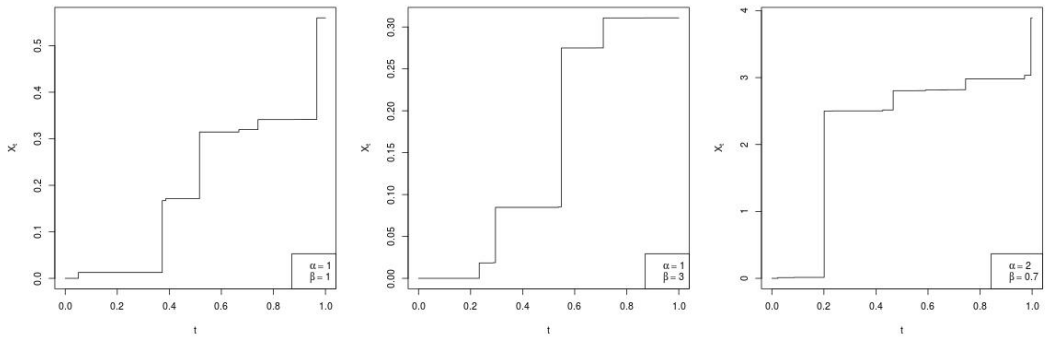

Figure 9: Bridge sample paths of Gamma processes.

As an experimental verification of the correctness of the Gamma bridge sampling algorithm, we run six test cases. In each case we use the algorithm with different input and sample 10^5 bridge points. From these we in each case create histograms and compare with the graph of the densities produced by equation (38). The test cases are as follows:

I $\tau_x = 0.5$, $\tau_y = 0.5$, $\alpha = 1$ and $z = 1$.

II Same as case I, but with $\alpha = 2$.

III Same as case I, but with $\alpha = 0.5$.

IV $\tau_x = 0.3$, $\tau_y = 0.7$, $\alpha = 1$ and $z = 1$.

V $\tau_x = 0.8$, $\tau_y = 0.2$, $\alpha = 1$ and $z = 1$.

VI $\tau_x = 0.6$, $\tau_y = 1.4$, $\alpha = 0.7$ and $z = 0.5$.

Graphical output of the numerical experiment can be seen in figure 10. In all six cases we see that the simulations and the theoretical bridge densities fits well. All the three first test cases are sampling bridge points at half way between the end points in time. This gives symmetric distributions as wee can see from the all the plots on the first row in figure 10. From this row we can also see that increasing (decreasing) α puts less (more) mass to the center of the bridge density. In the plot for case IV, we see that the symmetric get skewed towards small increments between the bridge point and the lower starting point. This is expected since we have moved the time of sampled bridge points towards the lower end point in time t_ℓ. The opposite effect for the skewness is observed in the plot for case V, where we have moved the time of sampled bridge points towards the upper end point in time t_ℓ. In the last test case VI, we have narrowed the total increment $Z = X_{t_u} - X_{t_\ell} = 0.5$ and we note that the support of the density is narrowed too. Here we have also chosen the time of the sampled bridge point a bit closer to t_ℓ and we can observe a similar effect for the skewness as in case IV.

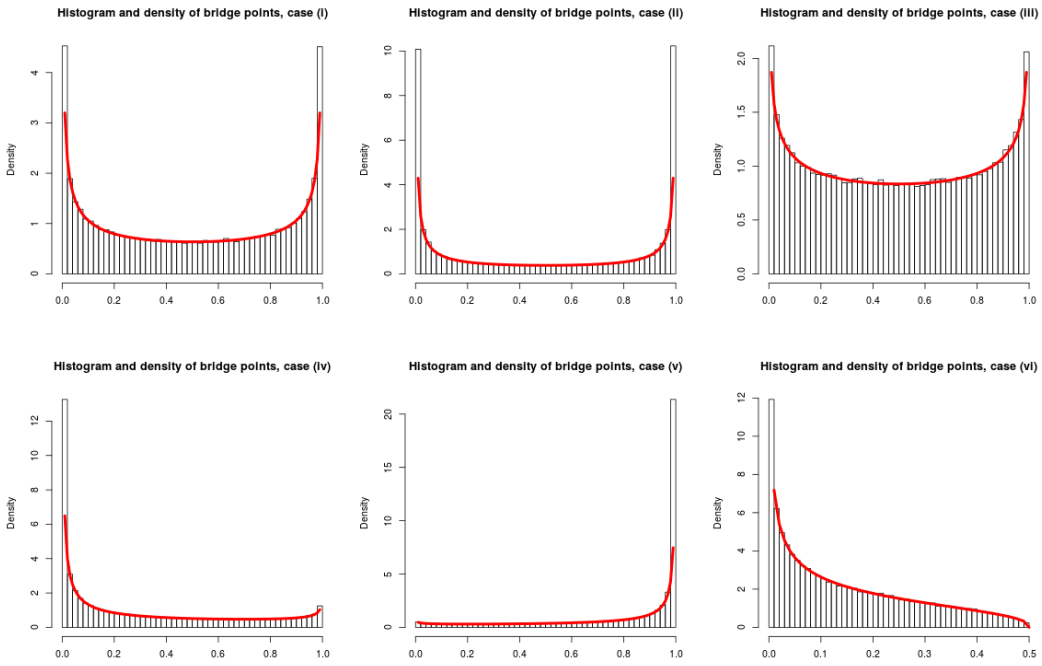

Figure 10: Histogram of simulated Gamma bridge points compared with their theoretical bridge distribution. The bridge distributions $f_{W|Z}(w)$ is in each case given by equation (38) and insertion of the parameters I-VI.

5.1.2 Inverse Gaussian bridge sampling

The source of this derivation is [21], but we also differ a bit from this paper. As mentioned above the construction of a bridge algorithm for the IG-process is more complicated than

for the Gamma process. We need to start out by presenting a preliminary result on generation of random variables from transformations with multiple roots, that is derived as a main result in [17]. Suppose a non-zero (except on a closed set) continuous function g is such that its derivative g' exists and that we have a relation between two continuous random variables Q and S on the form

$$Q = g(S), \tag{40}$$

where for a given realization $Q = q$, (40) have n distinct solutions $s_1, s_2, ..., s_n$. Moreover suppose $S \sim f_S(s)$ and that we are interested in generating realizations of S using the transformation (40). An arising question is then which one of the n roots of (40) should be chosen. The main result derived in [17] is that given $Q = q$, the ith root of (40) is a realization of S with probability

$$p_i(q) = \left(1 + \sum_{k=1, k\neq i}^{n} \left|\frac{g'(s_i)}{g'(s_k)}\right| \cdot \frac{f_S(s_k)}{f_S(s_i)}\right)^{-1}. \tag{41}$$

In practice we typically derive expressions for the roots $s_1, s_2, ..., s_k$ as a function of q. Note that in order to make use of (41), we must know the distributions of both S and Q.

Letting X_t, be the IG process, meaning $X_t - X_s \sim IG(\mu(t-s), \eta(t-s)^2)$, equation (37) becomes

$$f_{W|Z}(w) = \sqrt{\frac{\eta}{2\pi}} \left(\frac{wy}{z}\right)^{-3/2} \frac{\tau_w \tau_y}{\tau_z} \exp\left(-\frac{\eta}{2}\left(\frac{\tau_w^2}{w} + \frac{\tau_y^2}{y} - \frac{\tau_z^2}{z}\right)\right) \tag{42}$$

Define

$$Q = \eta\left(\frac{\tau_w^2}{W} + \frac{\tau_y^2}{Y} - \frac{\tau_z^2}{Z}\right), \tag{43}$$

which we can recognize from the exponent in (42). Q can also express as

$$Q = \frac{\eta\tau_y^2}{Z} \cdot \frac{\left(\frac{Y}{W} - \frac{\tau_y}{\tau_w}\right)^2}{\frac{Y}{W} \cdot \frac{\tau_y^2}{\tau_w^2}}.$$

If we now define $\lambda = \frac{\eta\tau_y^2}{Z}$, $\kappa = \frac{\tau_y}{\tau_w}$ and $S = \frac{Y}{W}$ we can write Q as a function of S only, given Z,

$$Q = \lambda\frac{(S-\kappa)^2}{S\kappa^2} := g(S). \tag{44}$$

Note that S must be strictly positive, from the fact that W and Y are increments of the strictly increasing Inverse Gaussian process. The Inverse Gaussian bridge sampling algorithm requires us to sample realizations of S, which can be done using the relation (44), as long as we can derive the distributions of both Q and S. This is a quadratic equation in S, where for any realization $Q = q$ we note that if s^* is a solution of (44), then $\frac{\kappa^2}{s^*}$ is the other solution since

$$g\left(\frac{\kappa^2}{s^*}\right) = \lambda\frac{\left(\frac{\kappa^2}{s^*} - \kappa\right)^2}{\frac{\kappa^2}{s^*}\kappa^2} = \lambda\frac{s^*\left(\frac{1}{s^*\kappa}\left(\kappa - s^*\right)\right)^2}{\kappa^4} = \lambda\frac{(s^*-\kappa)^2}{s^*\kappa^2} = g(s^*). \tag{45}$$

Solving $q = g(s)$ we therefore obtain the roots

$$s_1 = \kappa \left(1 + \frac{q\kappa - \sqrt{4\lambda q\kappa + q^2\kappa^2}}{2\lambda} \right) \quad \text{and} \quad s_2 = \frac{\kappa^2}{s_1}. \tag{46}$$

The probabilities for choosing either of the roots in (46) can be calculated using equation (41), since the function g in (44) is clearly is non-zero except on the closed set $s = \kappa$ and continuous (recall that $s > 0$). If we let $f_S(s)$ denote the density function of S, we need in our case to obtain a computable expression

$$p_1 = \left(1 + \left| \frac{g(s_1)}{g(s_2)} \right| \cdot \left| \frac{f_S(s_2)}{f_S(s_1)} \right| \right)^{-1}, \tag{47}$$

from which we also can compute $p_2 = 1 - p_1$. A routine computation gives us the ratio of derivatives as

$$\left| \frac{g'(s_1)}{g'(s_2)} \right| = \left(\frac{\kappa}{s_1} \right)^2.$$

Conditioning on Z, we see that $S = \frac{Y}{W} = \frac{z-W}{W}$, so that $\frac{ds}{dw} = -\frac{z}{w^2}$. Letting λ and κ be defined as above, we find distribution of S from (42)

$$\begin{aligned} f_S(s) = f_{W|Z}(w) \cdot \left| \frac{ds}{dw} \right|^{-1} &= \sqrt{\frac{\eta}{2\pi}} \frac{\tau_w \tau_y}{\tau_z} \left(\frac{wy}{z} \right)^{-3/2} e^{-\frac{\eta}{2}\left(\frac{\tau_w^2}{w} + \frac{\tau_y^2}{y} - \frac{\tau_z^2}{z} \right)} \cdot \frac{w^2}{z} \\ &= \sqrt{\frac{\eta\tau_y^2}{2\pi z}} \frac{1}{1 + \frac{\tau_y}{\tau_w}} \frac{z}{w} \left(\frac{y}{w} \right)^{-3/2} e^{-\frac{\eta}{2}\left(\frac{\tau_w^2}{w} + \frac{\tau_y^2}{y} - \frac{\tau_z^2}{z} \right)} \\ &= \sqrt{\frac{\lambda}{2\pi}} \frac{1}{1+\kappa} (1+s) s^{-3/2} e^{-\frac{\lambda(s-\kappa)^2}{2\kappa^2 s}}. \end{aligned} \tag{48}$$

By (45) we have that the exponential factors will cancel in the ratio we need to compute p_1 and we have that

$$\frac{f_S(s_2)}{f_S(s_1)} = \frac{s_1^2(\kappa^2 + s_1)}{\kappa^3(1 + s_1)}.$$

Finally, by insertion in (47), the probabilities of picking each of the roots p_1 and p_2 can be expressed as

$$p_1 = \frac{\kappa(1+s_1)}{(1+\kappa)(\kappa+s_1)} \quad \text{and} \quad p_2 = 1 - p_1. \tag{49}$$

Now, the only thing that is left is to obtain the distribution of Q. The authors of [20] applies a result called Tweedie's theorem, which can be found in [23] and adapt this to our problem. We will however, use the relation (44) and exploit that we have already obtained the distribution of S, which is a more intuitive way of deriving Q's distribution. Since the function $g(s)$ is not one-to-one, we have to take extra care. For a background on the distribution of transformations of random variables, where transformation is not one-to-one, see [4]. The main idea is to detect intervals where the transformation is one-to-one, before applying the standard transformation formula on each such interval and

in the end obtaining the desired distribution by summing over the results on each such interval. From equation (44), we can compute

$$g'(s) = \frac{\lambda}{\kappa^2}\left(1 - \frac{\kappa^2}{s^2}\right), \tag{50}$$

and we see that for $s > 0$, $g'(s) = 0$ if and only if $s = \kappa$. Moreover we see that $g'(s) < 0$ for $s < \kappa$ and $g'(s) > 0$ for $s > \kappa$, meaning that g is decreasing on $(0, \kappa)$ and increasing on (κ, ∞). If we let s_1 be the lower root of (43), it is clear that $0 < s_1 < \kappa$ and $\kappa < s_2 < \infty$ as long as they are distinct. If we define $g_1(s_1) = g(s_1)$ for $0 < s_1 < \kappa$ and $g_2(s_2) = g(s_2)$ for $\kappa < s_2 < \infty$, we have that g_1 and g_2 are both one-to-one. For the distribution of Q, we can therefore establish.

$$f_Q(q) = f_S(s_1)\left|\frac{dg_1^{-1}(q)}{dq}\right| + f_S(s_2)\left|\frac{dg_2^{-1}(q)}{dq}\right|. \tag{51}$$

Furthermore, for $i = 1, 2$,

$$\frac{dg_i^{-1}(q)}{dq} = \frac{ds_i}{dq} = \left(\frac{dq}{ds}\right)^{-1}_{s=s_i},$$

which we can compute from (50). Inserting these derivatives in (51) we obtain

$$\begin{aligned}
f_Q(q) &= \sum_{i=1}^{2} f_S(s_i)\left|\frac{dq}{ds}\right|^{-1}_{s=s_i} \\
&= \sum_{i=1}^{2} \sqrt{\frac{\lambda}{2\pi}}\frac{1}{1+\kappa} s_i^{-3/2}(1+s_i)e^{-q/2}\frac{\kappa^2}{\lambda}\frac{s_i^2}{s_i^2-\kappa^2} \\
&= \frac{1}{\sqrt{2\pi}}e^{-q}\frac{\kappa}{1+\kappa}\sum_{i=1}^{2}\underbrace{\frac{\kappa\sqrt{s_i}}{\sqrt{\lambda}(s_i-\kappa)}}_{=q^{-1/2}}\frac{1+s_i}{s_i+\kappa} \\
&= \left[\frac{1}{\sqrt{2\pi}}q^{-1/2}e^{-q/2}\right]\frac{\kappa}{1+\kappa}\left(\frac{s_1+1}{s_1+\kappa} + \frac{s_2+1}{s_2+\kappa}\right).
\end{aligned}$$

We can recognize the expression in the square brackets above as the density of chi-squared distribution with one degree of freedom (χ_1^2-distribution). Inserting the relation $s_2 = \frac{\kappa^2}{s_1}$, the factor outside the square brackets reduces to 1. Therefore we have that

$$f_Q(q) = \frac{1}{\sqrt{2\pi}}q^{-1/2}e^{-q/2}, \quad q \geq 0. \tag{52}$$

That is $Q \sim \chi_1^2$. Being able to sample S, we proceed in the same way as for the Gamma bridge by noting that $S = \frac{Y}{W} = \frac{X_{t_u} - X_t}{X_t - X_{t_\ell}}$, giving

$$X_t = X_{t_\ell} + \frac{X_{t_u} - X_{t_\ell}}{1+S}. \tag{53}$$

Note again the structure of (53). Since any realization $S = s$ will be strictly positive, $\frac{1}{1+s} \in (0, 1)$ and we start again with the initial value of the process and add a proportion

of the total increment of the process from t_ℓ to t_u. We now present a summary of the IG-bridge sampling algorithm.

Data: t,t_ℓ, t_u, X_{t_ℓ}, X_{t_u} and η
Result: $X_t|X_{t_\ell,X_{t_u}}$
$\lambda \leftarrow \frac{\eta\tau_y^2}{X_{t_u}-X_{t_\ell}}$;
$\kappa \leftarrow \frac{t_u-t}{t-t_\ell}$;
draw $Q \sim \chi_1^2$;
$S \leftarrow \kappa\left(1+\frac{Q\kappa-\sqrt{4\lambda Q\kappa+Q^2\kappa^2}}{2\lambda}\right)$;
draw $U \sim Uniform[0,1]$;
if $U > \frac{\kappa(1+S)}{(1+\kappa)(\kappa+S)}$ **then**
  $S \leftarrow \frac{\kappa^2}{S}$
end
$X_t \leftarrow X_{t_\ell} + \frac{X_{t_u}-X_{t_\ell}}{1+S}$ **return**

Algorithm 2: IG bridge sampling

Before examining the output of this algorithm, display some numerical evidence in figure 11, showing that the samples of S really comes from the distribution obtained in equation (48). Again we compare simulated realizations with the theoretical distributions for different values of the parameters κ and λ. We do this because the whole IG-bridge sampling algorithm relies on the ability to sample from the distribution of S. In our work with the Gamma process, the parameter P was analogous to S, but we where able to obtain a well known distribution for P in the Beta distribution, and then assumed a sampler for the Beta-distribution was available. For the parameter S, the obtained distribution are rather rare, and assuming an available sampler from it is not realistic. Hence, we created our own sampler.

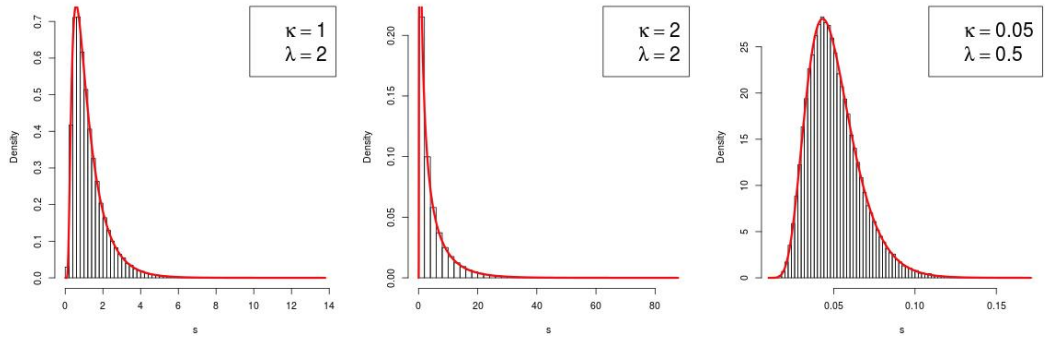

Figure 11: Histograms of simulated realizations of S, compared with theoretical distributions (48) (red lines) for different values of κ and λ.

Parameter values are indicated on each plot. We see that there is a good fit between the simulated samples and the densities of the theoretical distributions. Figure 12 displays sample paths of the IG-process simulated with the bridge sampling algorithm. These plots should be compared with the plots in figure 5 in section 3.2.2, since the parameters are

the same. Again we have done the simulations on $[t_\ell, t_u] = [0, 1]$ with endpoints $X_{t_\ell} = 0$ and $X_{t_u} \sim IG(\mu, \eta)$.

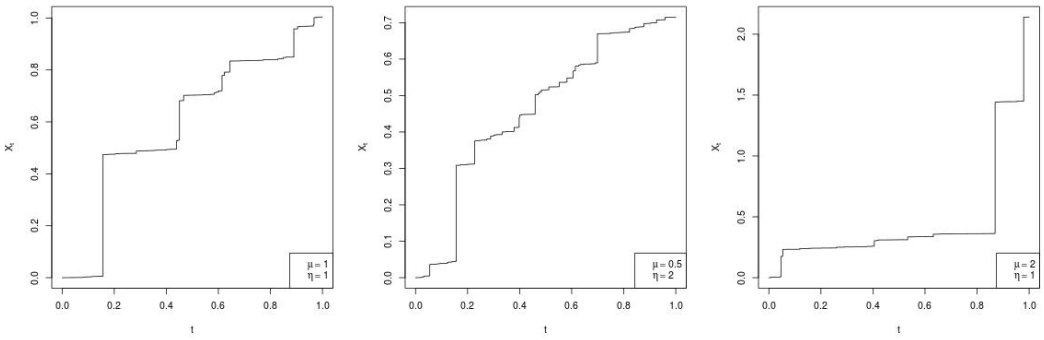

Figure 12: Bridge sample paths of IG processes.

Before leaving the IG-bridge sampling algorithm, we will also in this case set up some numerical experiments to verify its correctness. We use the same type of setup as with the Gamma bridge sampling algorithm and simulate 10^5 bridge points with six different parameter configurations. Then we compare histograms of the simulated output with the theoretical distribution, coming from equation (42), in each case. The six test cases are as follows:

I $\tau_x = 0.5$, $\tau_y = 0.5$, $\eta = 1$ and $z = 1$.

II Same as case I, but with $\eta = 0.5$.

III Same as case I, but with $\eta = 2$.

IV $\tau_x = 0.3$, $\tau_y = 0.7$, $\eta = 1$ and $z = 1$.

V $\tau_x = 0.8$, $\tau_y = 0.2$, $\eta = 1$ and $z = 1$.

VI $\tau_x = 0.6$, $\tau_y = 1.4$, $\eta = 0.7$ and $z = 0.5$.

Figure 13 shows the output of the experiment and we immediately see that the theoretical densities fits well to the histograms.

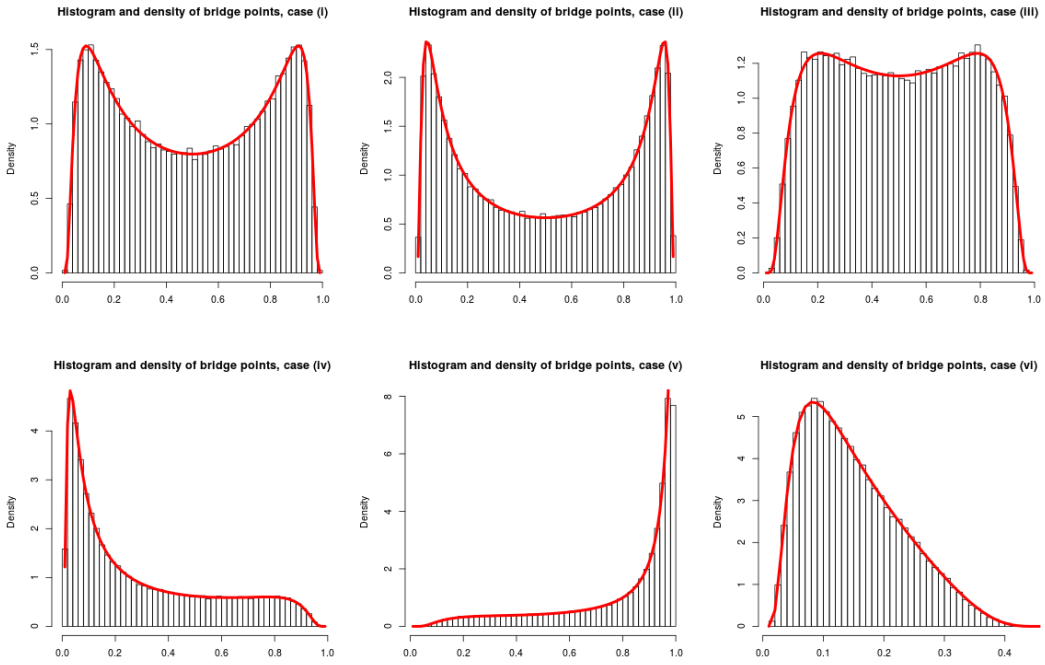

Figure 13: Histogram of simulated IG bridge points compared with their theoretical bridge distribution. The bridge distributions $f_{W|Z}(w)$ is in each case given by equation (42) and insertion of the parameters I-VI.

For the cases I, II and III, we first note that the distributions are symmetric, like we saw also in the Gamma case (figure 10). For the parameter η, we can see that decreasing (increasing) the parameter put less (more) mass to the center of the distribution for the bridge point increment. Moving t towards t_ℓ in case IV, we observe that the distribution becomes skewed towards lower increments. In case V, when we move t towards t_u, the opposite effect can be seen. This also expected and we saw the same effect in the Gamma experiments. For the last case VI, we have used a smaller total increment $z = 0.5$ and we note that the density gets concentrated on values on $(0, 0.5)$. The time t of the bridge point realization is moved a bit closer to t_ℓ, causing the distribution to become left-skewed. We also note there is a general difference between the Gamma bridge distributions and the IG bridge distributions, on how they look like. However many of the same effects are observed when changing parameters in the experiments.

5.1.3 Simulation of passage times using bridge sampling algorithms

Bridge sampling algorithms can also be used to simulate first passage times of stochastic processes. In our applications, these first passage time will be lifetime T_L, as defined by equation (14), that is the first time at which a stochastic process $X = \{X_t, t \geq 0\}$ with continuous state space passes a certain level L. We will here assume that X is a subordinator. Suppose again we have observed the process at two time points t_ℓ and t_u. Further, if $X_{T_\ell} < L$ and $X_{t_u} > L$, we know that $t_\ell < T_L < t_u$. If the distribution of T_L is hard to obtain or hard to sample from, a bridge sampling algorithm may be applied to simulate realizations of T_L. Let *genericBridge* be a generic bridge sampling algorithm,

returning a bridge point X_t, taking input $X_{t_\ell}, X_{t_u}, t, t_\ell, t_u$ and θ, where $t_\ell < t < t_u$ and θ is a generic parameter vector specifying the generic subordinator X. The algorithm then goes as follows

Data: L, $\tilde{t}$, t_ℓ, t_u, X_{t_ℓ}, X_{t_u} and θ
Result: T
while $|t_u - t_\ell| > \tilde{t}$ **do**
  $\tilde{X} \leftarrow genericBridge(X_{t_\ell}, X_{t_u}, \frac{t_u+t_\ell}{2}, t_\ell, t_u, \theta)$;
  if $\tilde{X} > L$ **then**
    $X_{t_u} \leftarrow \tilde{X}$;
    $t_u \leftarrow \frac{t_u+t_\ell}{2}$
  end
  else
    $X_{t_\ell} \leftarrow \tilde{X}$;
    $t_\ell \leftarrow \frac{t_u+t_\ell}{2}$
  end
end
$T \leftarrow \frac{t_u+t_\ell}{2}$;
return T

Algorithm 3: Bisection algorithm for passage times.

This algorithm, that we will refer to as *bisectionAlgorithm* later, is based on the inequality $t_\ell < T_L < t_u$, continuously making the possible range of T_L-values narrower as long as the time distance $t_u - t_\ell$ is larger than some threshold $\tilde{t}$, before it returns T_L as the mean of the bounds. It uses a bisection approach, by computing the bridge point $\tilde{X}$ at the time half way between the bounds at each iteration. If $\tilde{X}$ is above L we update the upper value X_{t_u} and if $\tilde{X} < L$ we update the lower value X_{t_ℓ}, and simulate a new bridge point from the updated input. Figure 14 illustrates how the algorithm works, where the range of possible T_L-values is marked as a red interval on the time axis.

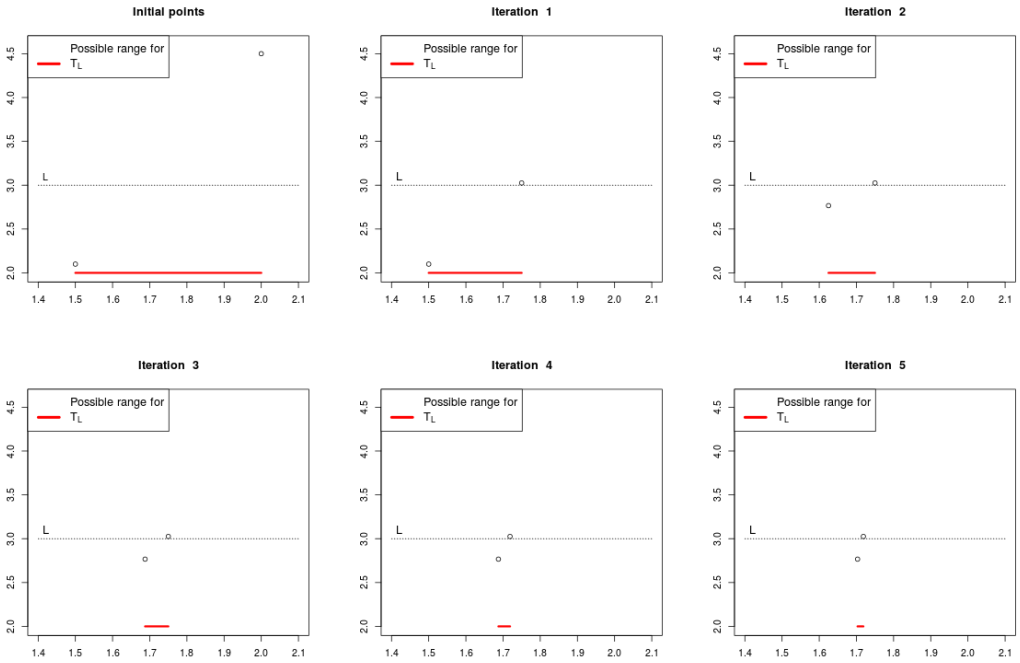

Figure 14: Illustration of algorithm 3.

The illustration shows a Gamma process, with $\alpha = 1$, observed at times $t_\ell = 1.5$ with $X_{t_\ell} = 2$ and $t_u = 2$ with $X_{t_u} = 4.5$. We can see that in the first iteration X_{t_u} is updated, before X_{t_ℓ} is updated in the second iteration an so on. As seen from the plots, the bisection approach ensures very fast convergence and the range of possible T-values is very limited already in the fifth iteration.

5.2 Simulation of the degradation model

5.2.1 MCMC's with continuous state space

This is a short summary of some key concepts, stated in [10], regarding convergence for a Markov Chain Monte Carlo (MCMC) sampler. Typically, MCMC-techniques are used in order to explore properties of some probability distribution π, that is hard to sample from with standard methods. The idea of this approach is to construct a discrete time Markov chain $X = \{X_n, n \geq 0\}$, that has π as its stationary distribution. As a remark, we have that in our case, the chain already as a part of the model, namely the Y-process, so it doesn't need to be constructed. Having established such a chain, one simulates it for a sufficient amount of time and estimate quantities of interest as sample path averages of X. By a sufficient amount of time, we mean two things. First the simulation must run long enough for the chain to converge to its stationary distribution. To ensure this, we define a sufficiently large burn-in period and discard all samples collected in this period. Secondly we need to collect enough samples from the stationary distribution, so that we can use the law of large numbers for these sample path averages. Let f represent a quantity of

interest and define the sample path average as.

$$\bar{f}_n = \frac{1}{n+1} \sum_{i=0}^{n} f(X_n), \tag{54}$$

We then need to ensure that

$$f_n \to \mathrm{E}_\pi[f(X)].$$

For a continuous state Markov chain, the following theorem is stated by [10]:

Theorem 3. *Suppose $X = \{X_n, n \geq 0\}$ is an irreducible Markov chain with continuous state space E and stationary distribution π, and let f be a real-valued function on E, such that $\mathrm{E}_\pi[|f(X)|] < \infty$. Then $P\left(\bar{f}_n \to \mathrm{E}_\pi[f(X)] \,|X_0 = x\right) = 1$ for π-almost all x, where $\bar{f}_n$ is given by equation (54).*

Since we know from earlier discussions that Y satisfies this theorem, we now want to construct an algorithm for simulating the Y-process and compute averages on the form (54).

5.2.2 The algorithm

We now present an algorithm for simulating the original model that was presented in section 4.1. Note that we could get more information out of this simulation, than what we have chosen to return in this algorithm. Most of the variables involved should be familiar from earlier sections, but some are new and will now be explained:

- N, number of samples of the Y-process collected after burn-in.
- $\mathbf{Y}$ - vector containing the history for the Y-process. That is $\mathbf{Y} = [Y_0, Y_1, ..., Y_N]$.
- $\mathbf{T}$ - vector of semi-regeneration times. That is $\mathbf{T} = [T_0, T_1, ..., T_N]$.
- $\mathbf{\Delta T}$ - vector of intervals between semi-regeneration times. That is $\mathbf{\Delta T} = [T_1 - T_0, T_2 - T_1, ..., T_N - T_{N-1}] := [\Delta T_1, \Delta T_2, ..., \Delta T_N]$.
- k - Number of corrective maintenance actions.
- l - Number of preventive maintenance actions.

A pseudo-code for the algorithm is given below and will be followed by an explanation of the most important parts of it.

Data: N,α,β,A,B,M,L,C_i,C_p,C_c and C_d
Result: $\widehat{EC_\infty}$ and/or other desired estimates.
initialize $\mathbf{\Delta T}$, $\mathbf{T}$, with $T_0 = 0$ and $\mathbf{Y}$, with $Y_0 \sim \pi$.;
$k \leftarrow 0$;
$l \leftarrow 0$;
$d_t \leftarrow 0$;
for $i \leftarrow 0$ **to** $N-1$ **do**
  $\Delta T_{i+1} \leftarrow m(Y_i)$;
  $T_{i+1} \leftarrow T_i + \Delta T_i$;
  draw $G \sim Ga(\alpha \cdot \Delta T_i, \beta)$;
  $Y_{i+1} \leftarrow Y_i + G$;
  if $Y_{i+1} \geq L$ **then**
    $k \leftarrow k+1$;
    $t_\ell \leftarrow T_i$;
    $t_u \leftarrow T_{i+1}$;
    $X_{t_\ell} \leftarrow Y_i$;
    $X_{t_u} \leftarrow Y_{i+1}$;
    run $bisectionAlgorithm(L, \tilde{t}, t_\ell, t_u, X_{t_\ell}, X_{t_u}, \alpha, \beta)$, with desired threshold $\tilde{t}$, to generate T_L;
    $d_t \leftarrow d_t + (T_{i+1} - T_L)$;
  end
  else if $Y_{i+1} \geq M$ **then**
    $l \leftarrow l+1$;
    $Y_{i+1} \leftarrow 0$;
  end
end
$\widehat{E_\pi[T]} \leftarrow \frac{\sum_{i=1}^{N-1} \Delta T_i}{N-1}$;
$\widehat{E_\pi[N_T^{(p)}]} \leftarrow \frac{l}{N-1}$;
$\widehat{E_\pi[N_T^{(c)}]} \leftarrow \frac{k}{N-1}$;
$\widehat{E_\pi[d_T]} \leftarrow \frac{d_t}{N-1}$;
$\widehat{EC_\infty} \leftarrow C_i \cdot \frac{1}{\widehat{E_\pi[T]}} + C_p \cdot \frac{\widehat{E_\pi[N_T^{(p)}]}}{\widehat{E_\pi[T]}} + C_c \cdot \frac{\widehat{E_\pi[N_T^{(c)}]}}{\widehat{E_\pi[T]}} + C_d \cdot \frac{\widehat{E_\pi[d_T]}}{\widehat{E_\pi[T]}}$;
return $\widehat{EC_\infty}$

Algorithm 4: Algorithm for the original degradation and maintenance model.

This algorithm has a main routine that happens inside the for-loop. To ensure that the starting value $Y_0 \sim \pi$, we must run this routine sequentially in for a burn-in period, before we start. The main routine starts out by setting a date for the next inspection. Then the next value of the Y-process is generated and we are ready for inspection. If we detect a failure, the counter k for corrective maintenance actions is incremented, before we simulate the time T_L for when the failure actually occurred. Having simulated this time, we have that the simulated downtime because of the failure is $(T_{i+1} - T_L)$, that is the failure time subtracted from the inspection time, and add it to the total accumulated downtime d_t. In the end the value of the Y-process is set to 0, in accordance with

equation (24) in section 4, representing a perfect repair. Similarly, if the outcome of the inspection is preventive maintenance, l is incremented and the Y-process is reset to 0. After having collected the desired number of samples, one can obtain estimates of many different quantities in the model. Our choice of returning the expected cost rate, is mostly because this is the quantity we have described the most in this text. Having this algorithm as a point of departure, it is easy to implement the extensions we have discussed and possibly many others. For example can the non-perfect inspection extension from section 4.2.2 be included by flipping a coin with probabilities q and $1-q$ in the "else if $M \geq L$"-condition.

Algorithm 4 will also be the basis for implementation of the extensions we proposed in the end of section 4.

6 Numerical experiments

We now present a series of numerical experiments, using the algorithms from last sections. Implementation of the algorithms is an easy task in standard programming languages and here we have used the statistical programming environment R. We start out by repeating an experiment from [11], where we also try to play around a little bit with the decision variables A, B and M. Then we run a simulation where we replace the Gamma degradation model with an Inverse Gaussian degradation model. Here we also try to compare the two processes, with choice of similar parameters and we make an extra check of the bridge sampler for the IG-process. In the last experiment, we implement the model with non-perfect inspections and examines some effects from this. Also here we try to change on A, B and M. The goal of the experiments is mainly to study how the models behaves and whether or not this behavior makes sense. This will be done both by computing various estimates and by looking at relevant plots.

6.1 Replicating an old experiment

To verify that the algorithm 4 performs correctly, we replicate an old experiment given in [11]. We also study some effects of changes in the decision variables.

6.1.1 Reproducing results

In this expeiment, the fixed parameters are as follows:

$$C_i = 25, \quad C_p = 50, \quad C_c = 100, \quad C_d = 250, \quad \alpha = 1, \quad \beta = 1, \quad L = 12. \tag{55}$$

By numerical optimization, optimal values for the decision variables are found in [11] to be

$$A = 5.5, \quad B = 9, \quad M = 5.6,$$

with and long term expected cost rate computed to be

$$EC_\infty = 12.23.$$

Running algorithm 4, with $N = 3 \cdot 10^5$ we are able to reproduce this result with $\widehat{EC_\infty} = 12.23$. This is a very strong indicator for that our code performs correctly. It should be mentioned that using this MCMC approach we need a large number of iterations to get the desired accuracy and that we obtain fluctuating result for smaller experiments. However the results are reasonable already for much smaller experiments. Figure 15 shows a trace plot of $\widehat{EC_\infty}$ computed in every iteration.

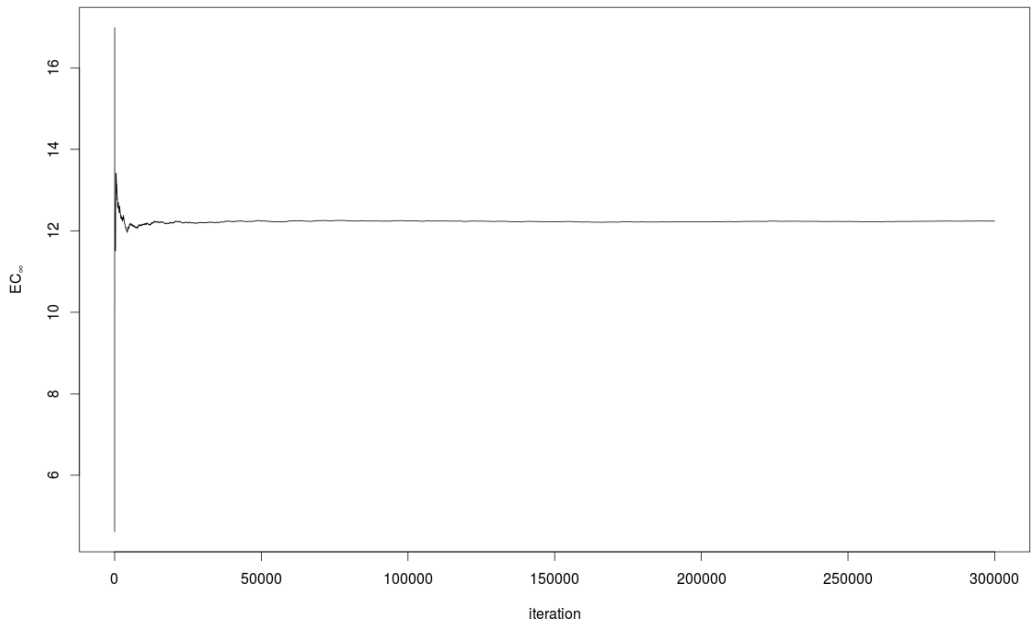

Figure 15: Trace plot of $\widehat{EC_\infty}$

We see from this plot that after a little more than 50 000 iterations, the fluctuations gets smaller. To get an impression of the fluctuations, we may for example for experiments with 60 000 iterations produce for example 12.23, 12.26, 12.21, 12.28 as estimates for EC_∞. Even though, this is not very accurate numbers, the results are not too bad as they are all in the neighborhood of 12.23.

6.1.2 A closer look at the density of the stationary distribution

We will now use the same values as in (55) for the fixed parameters, and try different values for the decision variables A, B and M. The purpose of this experiment is to study how the density $\pi(y)$ of the stationary distribution for Y, behaves and try to interpret the results. The best we can do for examine the behavior of π is to look at histograms from simulated samples of Y. Recall that according to equation (30), we expect $\pi(y)$ to be a linear combination of a Dirac delta at the regeneration set $\{0\}$ and some other density having positive support. We will now introduce four experiments, specified by the choices of decision variables:

I $A = 5.5$, $B = 9$, $M = 5.6$. Remark: same as in section 6.1.1

II $A = 0$, $M = 0$. Remark: Gives eqiudistant inter-inspection times.

III $A = 5.5$, $B = 9$, $M = 11$. Remark: M moved closer to L.

IV $A = 5.5$, $B = 0$, $M = 2$. Remark: M moved closer to 0.

Figure 16 displays the graphical output of the four experiments, along with the estimated value of EC_∞. Note that the vertical axis is densities, not probabilities.

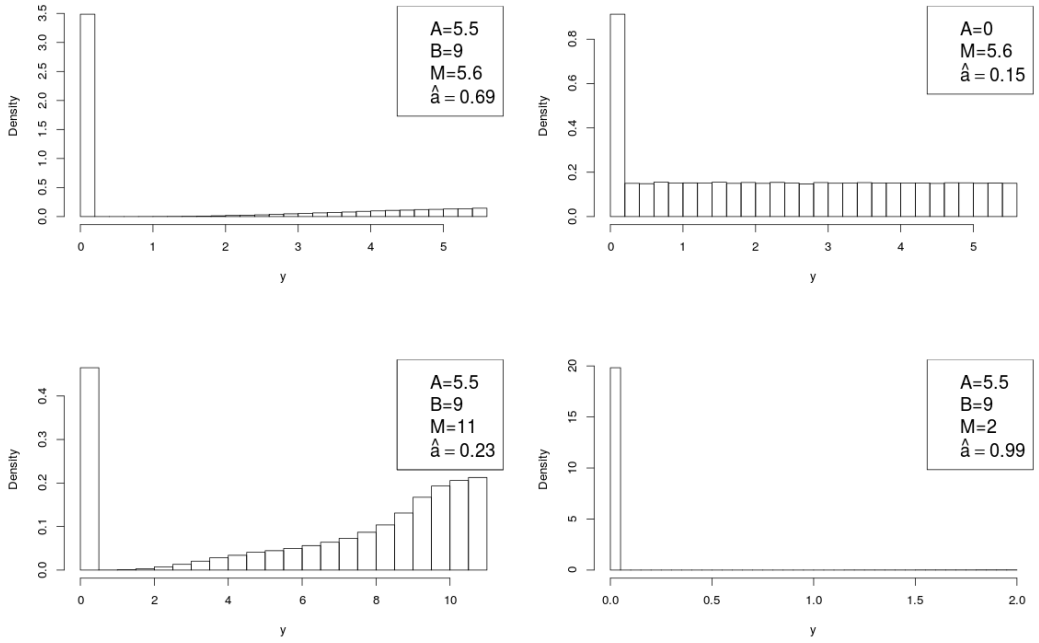

Figure 16: Histograms of simulated samples of the process Y.

In all the cases, the column in the origin represents mainly observations of the process in state 0 (only in case II it is influenced by a small amount of neighboring observations). We have also included the cumputed estimate $\hat{a}$ of a, the stationary probability of being in the regeneration set $\{0\}$. $\hat{a}$ is just the number of observations of the Y-process in 0, divided by the number of observations. Starting with experiment I, the histogram shows that the origin dominates and that density on $y > 0$ seems to increasing towards larger y. As explained in the previous section, the choices for A, B and M are found to be optimal in this case. We can provide some more numbers from the experiment to gain insight on the behavior of π. For case I, we have that

$$\frac{C_i}{\widehat{E_\pi[T]}} = 4.37, \quad \frac{C_p \cdot \widehat{E_\pi[N_T^{(p)}]}}{\widehat{E_\pi[T]}} = 5.79, \quad \frac{C_c \cdot \widehat{E_\pi[N_T^{(c)}]}}{\widehat{E_\pi[T]}} = 0.61, \quad \frac{C_d \cdot \widehat{E_\pi[d_T]}}{\widehat{E_\pi[T]}} = 1.44.$$

This is somewhat the ideal combination of cost distribution given the cost rates, and we can learn more about comparing the corresponding outputs from the other experiments. In case II, the numbers are:

$$\frac{C_i}{\widehat{E_\pi[T]}} = 25, \quad \frac{C_p \cdot \widehat{E_\pi[N_T^{(p)}]}}{\widehat{E_\pi[T]}} = 7.55, \quad \frac{C_c \cdot \widehat{E_\pi[N_T^{(c)}]}}{\widehat{E_\pi[T]}} = 0.02, \quad \frac{C_d \cdot \widehat{E_\pi[d_T]}}{\widehat{E_\pi[T]}} = 0.02.$$

Here the expected cost rate is 25, because setting $A = 0$ leads to sequential inspections with the time between inspections being 1. This is the shortest inter-inspection time possible to obtain in this model, when we use the function m, given by (25) in section 4. Then it is also reasonable that this approach leads to a lot of preventive maintenance actions, which we also can see from the estimate of the expected long term cost rate. The

costs for corrective maintenance and downtime are in this case negligible. From figure 16 we see that the density for $y > 0$ looks uniform. A similar response with larger inter-inspection times could also be constructed by letting $B \to \infty$. In that case $1 + A$ would be the time between inspections. In other words, a pure time based inspection policy is contained in the model. This policy performs good if inspections are cheap, but as we can conclude from case II, the long inspection cost rate dominates and the method is expensive compared with case I. We proceed with the experiment III, where we have the estimates

$$\frac{C_i}{\widehat{E_\pi[T]}} = 8.90, \quad \frac{C_p \cdot \widehat{E_\pi[N_T^{(p)}]}}{\widehat{E_\pi[T]}} = 2.53, \quad \frac{C_c \cdot \widehat{E_\pi[N_T^{(c)}]}}{\widehat{E_\pi[T]}} = 3.21, \quad \frac{C_d \cdot \widehat{E_\pi[d_T]}}{\widehat{E_\pi[T]}} = 4.01.$$

Case III performs better than case II, and of course worse than the optimal case I. Compared with case I, we see from the histograms that the density seems to take the same increasing form as in case I, for $y > 0$, but it puts more mass to this region. Looking at the estimates, we see a that the costs are lowered for preventive maintenance actions, as we should expect when performing such tasks only when the item is about to fail. The price is paid on all the other areas. The corrective maintenance and downtime costs have not surprisingly increased and we also note that the cost rate for inspections are approximately doubled. The last experiment IV, performs pretty good and we see that the estimated global cost rate is not too far away from case I. Estimates computed here are

$$\frac{C_i}{\widehat{E_\pi[T]}} = 3.85, \quad \frac{C_p \cdot \widehat{E_\pi[N_T^{(p)}]}}{\widehat{E_\pi[T]}} = 7.39, \quad \frac{C_c \cdot \widehat{E_\pi[N_T^{(c)}]}}{\widehat{E_\pi[T]}} = 0.47, \quad \frac{C_d \cdot \widehat{E_\pi[d_T]}}{\widehat{E_\pi[T]}} = 1.21.$$

In addtion, we here have that $\widehat{E_\pi[T]} = 6.49$, which is very close to $m(0) = 6.5$. Letting $M \to 0$ would actually mean a pure clock-based maintenance policy, which is also contained in our model. From the numbers we see that almost every action is a preventive maintenance action. From histograms we seems that all mass is in the origin. This means that almost at any inspection a maintenance action takes place.

Other experiments have also been carried out and all of then leads to results that can be physically interpreted in the same fashion as we did with case I-IV. This means that the system behaves and responds in a reasonable way, taking its purpose and the model into account.

6.2 Simulations with extensions

We also include some examples where we simulate the models with extensions from section 4.2.1 and 4.2.2. Implementing the changes is straight forward, when using algorithm 4 as a basis. Through these experiments we also try to explore some properties and compare the results with the original model.

6.2.1 Using the IG-process as a degradation model

In another experiment in [11] using Gamma degradation with parameters

$C_i = 2, \quad C_p = 90, \quad C_c = 100, \quad C_d = 100, \quad \alpha = 1, \quad \beta = 0.2, \quad L = 60$,the optimal desicion variables are found to be

$$A = 4.4, \quad B = 45, \quad M = 50.$$

The computed long term cost rate is found to be $EC_\infty = 9.48$. We can replicate these results with our simulation framework. However we will now try to use an Inverse Gaussian degradation process, where we chose the parameters μ and η so that the mean increment and variance becomes the same as in the Gamma process case. Using the equations (18) and (22) from section 3, we find that we must have

$$\mu = 5 \quad \text{and} \quad \eta = 5.$$

We run the algorithm 4, where we substitute the Gamma and Gamma bridge sample with its Inverse Gaussian equivalents and obtain the estimate $\widehat{EC_\infty} = 9.83$. This is a little bit higher than in the Gamma case, but still it is certainly in the same neighborhood. Since we have chosen the parameters such that the mean and the variance are equal in the two degradation models, this is a reasonable result. The processes are however very different when we look at them as we can see from figure 17. Even though the processes produce a similar long term cost rate, there could by reason like for example the physics of the degradation model to prefer one over the other.

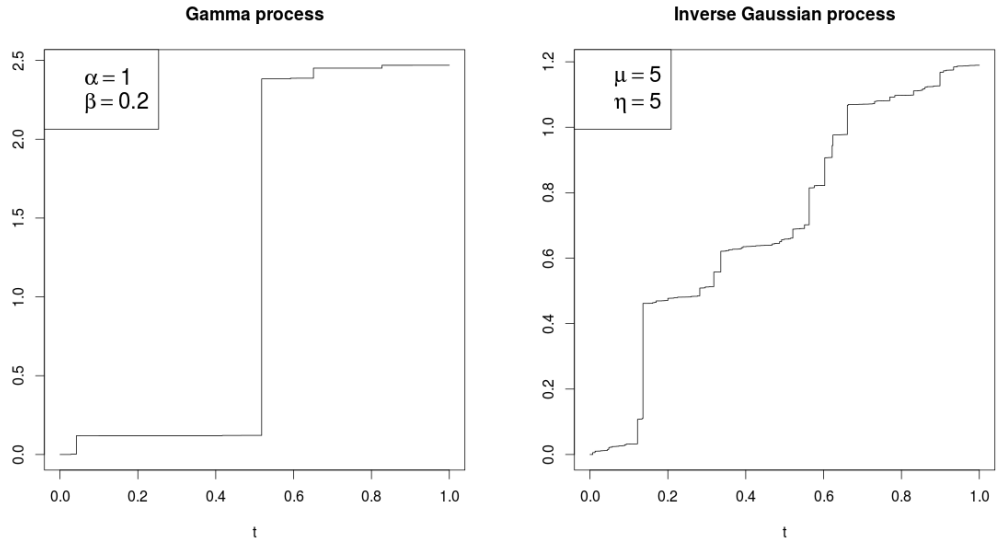

Figure 17: Gamma process with $\alpha = 1$ and $\beta = 0.2$ and the IGprocess with $\mu = 5$ and $\eta = 5$

In figure 18 we display histograms of observations of the Y process, in the same way as in section 6.1.2.

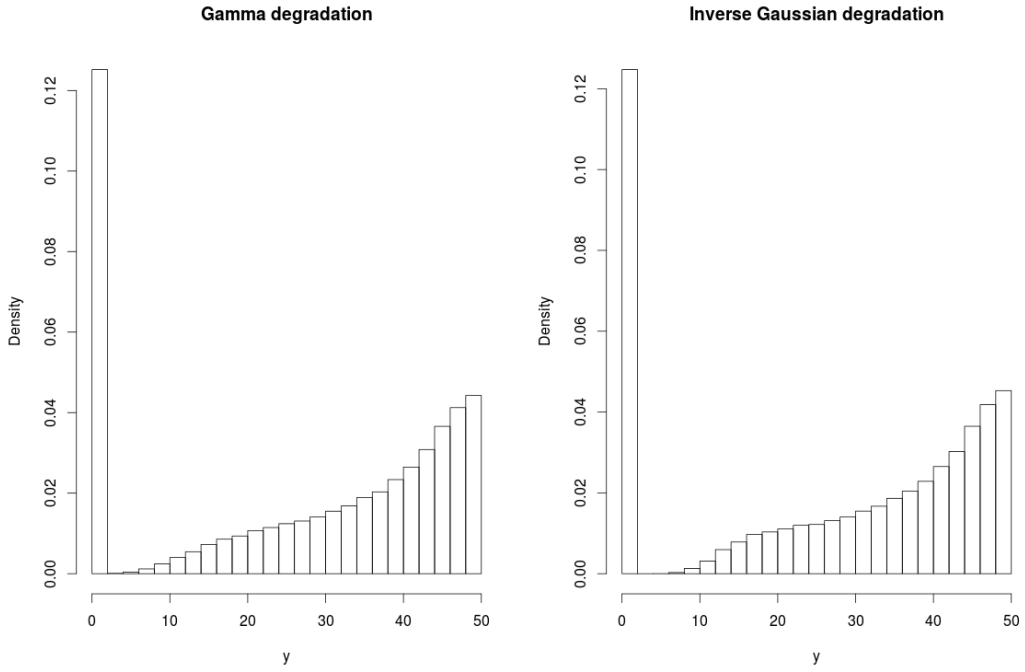

Figure 18: Histograms of observations of the Y-provess using the Gamma process with parameters $\alpha = 1$ and $\beta = 0.2$ to the left, and the IG process with $\mu = 5$ and $\eta = 5$ to the right.

We see that even if typical sample paths of the processes looks different, the stationary distribution π seems to behave more similar. Perhaps we can see a small difference in the shape of the histograms for $y > 0$, but the density at the the regeneration set is almost identical. When working with the IG-process we can also make an extra demonstration of the correctness of the IG bridge sampler. By letting all parameters be unchanged in the experiment above, except from letting $M = L$, we produce a "run to failure"-regime experiment with IG-degradation. Collecting samples of the passage times and subtract this from the last regeneration time, we can compare the empirical cdf of these, with the theoretical expression given by equation (23). To be precise, we will under the absence of preventive maintenance collect variables $T_{Ln} - S_n$, for $n = 0, 1, ...$, where T_{Ln} is the first time after S_n the process passes the level L, and check if they follow the distribution for the passage time of IG-process. Algorithm 3 is producing the realizations of T_{Ln}.

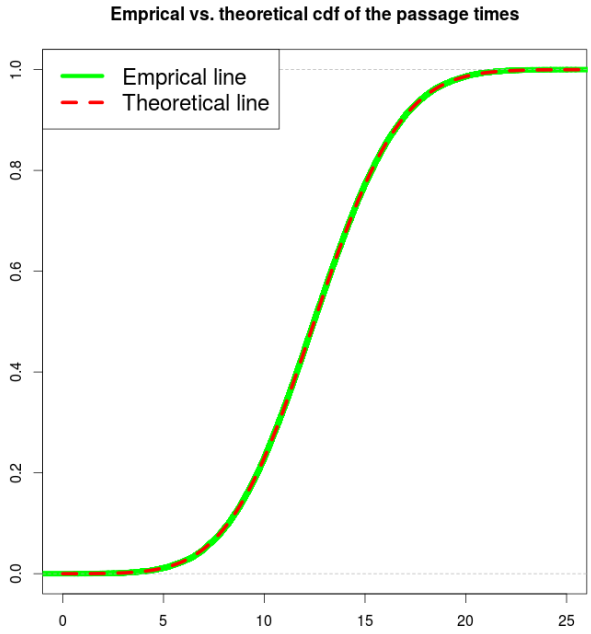

Figure 19: Comparing the simulated passage times with its theoretical distribution.

There is a very good fit between the theoretical and empirical pdf, so we see that the IG bridge sampler is working as intended.

6.2.2 Non-perfect inspections

We now run an experiment with non-perfect inspections as proposed in section 4.2.2. That is, when the process X is beyond M a signal telling us to do preventive maintenance is emitted, and we detect this signal with probability q as inspection. If the process reaches L, we will with probability one detect the failure. We use the Gamma process as degradation model and use same parameter configuration as in section 6.1.1. We run 11 experiments, where we start with $q = 1$, before we successively subtract 0.1 from q in the next experiment, ending with $q = 0$. The first experiment is thus another replication of section 6.1.1, meaning perfect inspections, while the last experiment is a "run to failure"-regime. In each experiment we compute all the estimated cost rates. In figure 20 we show how the different estimated cost rates changes in each of the simulations as we lower the detection probability q.

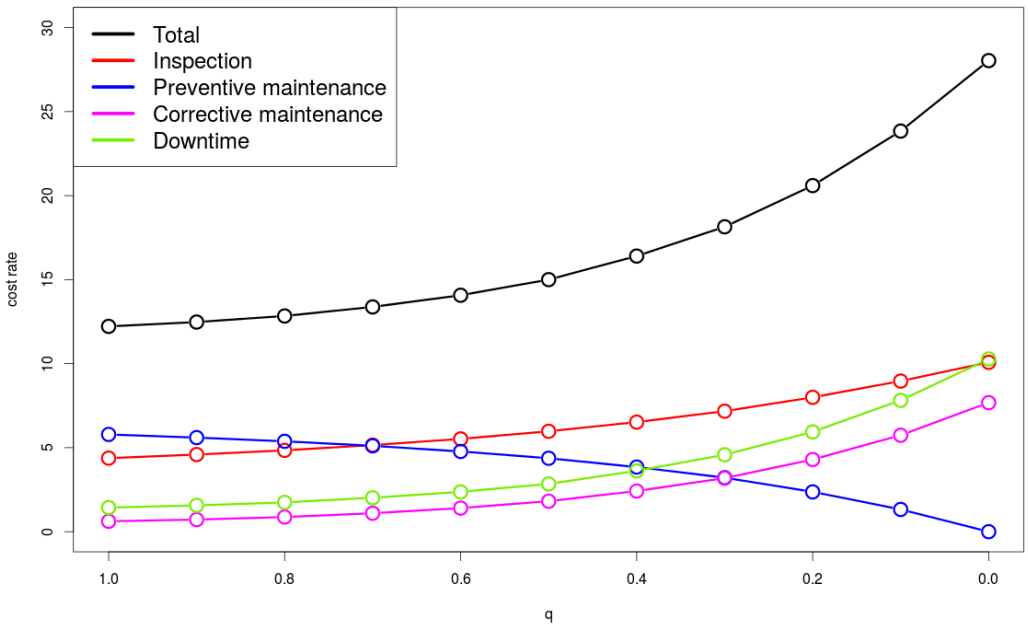

Figure 20: The estimated cost rates as a function of q. Note that q varies from 1 to 0 on the vertical axis.

The leftmost points are the same estimates we had in section 6.1.1. As we move towards smaller q, we see a small change in the beginning before the rates changes increases as q gets closer to 0. All the cost rates increases, except from the cost rate of preventive maintenance that decreases to 0 in the end. This is exactly what we would have expected. Of course the preventive maintenance cost will disappear, when we let the item run until failure. But we also see clearly the effect that preventive maintenance has on the costs, as both the correction and downtime costs changes a lot more than the inspection costs. In the case of $q = 1$, the optimal decision parameters where found by [11] to be $A = 5.5$, $B = 9$ and $M = 5.6$, with $EC_\infty = 12.23$. By letting q change, these parameters will no longer be optimal. If we change the decision variables one at the time, we will in this case see that the optimum doesn't change by very much. As an example, we have with $q = 0.5$, changed one decision variable at the time to get an impression of how the estimated cost rate EC_∞ changes. For all the parameters, we changed them between 1 and 10, with 1 as the step-length. The result is shown in figure 21.

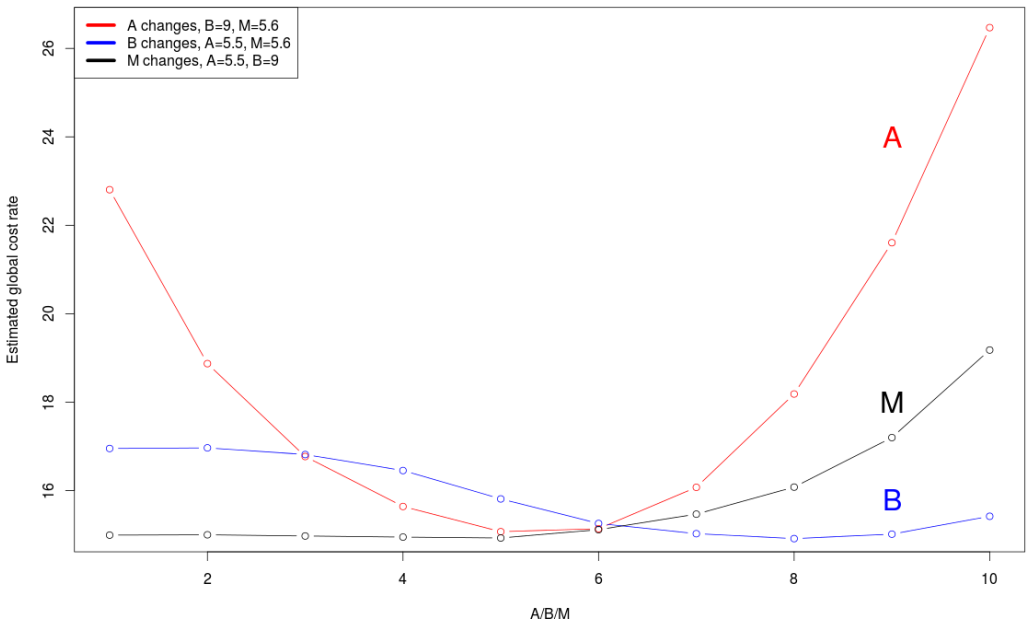

Figure 21: Response on the estimated global cost rate, while changing the decision variables one at the time with $q = 0.5$.

The red curve shows that $A = 5$ gives the lowest estimated cost, when $B = 9$ and $M = 5.6$ is fixed. The optimal value with $q = 1$ was $A = 5.5$. Along the blue curve we see that optimum is reached for $B = 8$, when $A = 5.5$ and $B = 9$. Here the original optimum was $B = 9$. Finally the along the black curve, which is a bit difficult to determine from the figure, the lowest computed estimate was at $M = 5$, with $A = 5.5$ and $B = 9$. Optimum with $q = 1$ was at $M = 5.6$. We also not the the estimated cost seems to be much more sensitive to changes in A than to changes in B or M.

7 Discussion

7.1 Adequacy of the model

The model we have studied here is constructed and grounded, with respect to both qualitative and statistical theory in reliability analysis. Under the assumption that an item's state can be summarized through one condition variable, and that we are able to observe this condition variable, we have seen from the simulations in section 6, that this models performs reasonable. We were able to explain and many of the observations we did, with referring to principles and theory at which the model is grounded. In [11], it is done additional experiments with change in the for example the costs, for example to see the effect of very cheap inspections. These experiments also provided results with intuitive explanations. The underlying assumption regarding the item's state however, is more dubious and will be addressed in the next section.

7.2 Recommendations for future work on the model

As we have discussed so far, there are a lot of things that could be done to improve on this model and in this section we will go through a couple of suggestions. An natural thing to start with is to find estimators for the parameters in the degradation models. We have not focused on parameter estimation, but the degradation processes discussed here should be well suited for both an frequentistic or Bayesian estimation framework. The papers [18] and [26] are both good places to start out in this case. Here we may also find examples of how to extend the models by letting the parameters be functions or random variables, which is a possible extension that was mentioned in section 3.2. [26] discusses in particular how such models may cause the IG-process to become non-Markovian, which again may cause many of our assumptions to no longer being valid. In other words, when incorporating new elements to the model, caution should be taken. A last remark on the degradation modeling is that we may try to use other càdlàg processes for the deterioration. The choice of using the IG-process was highly motivated the papers [26] and [25], and the main point of the numerical experiment in section 6.2.1 was to show that other models can make sense.

To be able to use these models in practice, we also need a way to measure a system's state variable. We have assumed that such measurements are always possible, but having discussed these models with reliability engineers, it is clear that they see this as a weakness in the model. How and what to measure is a problem that mainly should be addressed to engineers and other experts in the possible fields of application. [9] mentions erosion of dikes as an example of when deterioration can be described by a Gamma process. However it is clear to we can measure time series of a lot of variables that can be assumed to be correlated with the true condition of a system. Therefore it should be possible to investigate these correlations and exploit these to obtain a clearer impression of the true system state. A suggestion for a an approach here is to develop classification algorithms.

Another thing to consider is that if a model like this should be implemented in some business, the users should have the opportunity to constantly add data to the model. The way this model is constructed, this is definitely possible. On the other hand if the amount of data should become very large and occur very rapidly, the need for smart data handling will become present. This could for example be the case if we where continuously

monitoring the system and wanted to make real time estimates of some of the quantities. Different ways that data might flow into a model is explained in [13] and there is a lot of work being done on how to handle such data streams.

In this text, we only looked at a single-unit system. Many common systems in reliability analysis consists of more than one system, and more work could therefore be done on how to incorporate models like this in a multi-unit system. [5] is an example of how this could be done, where a similar model to ours is used for a two-component series system.

One can also look more into detailed assumptions in the model. As an example, what happens if we do not longer have perfect repair. One way to do this can be to say that the process may not be restored to the "good as new"-state, but with absolute certainty to some state below a threshold level. Then the regeneration set would not longer be just one point. Another way can be to say that the repair will take some random or fixed amout of time. Then downtime cost would also be associated with preventive maintenance actions. These two ideas could also be combined, or one could look closer into other details of the model.

Finally, we remember that in [11] and [9], numerical optimization of the decision variables was done. Our simulations are build on a more or less plain Monte Carlo method and optimization with this models would be very computer intensive. There are most likely ways to improve on the efficiency of the algorithm and develop optimization procedures. Through many of the numerical experiments, we also observed that to get very sharp estimates, the simulations had to go on for a very long time. This should also be possible to improve on.

7.3 Concluding remarks

Based on its construction and the numerical experiment, the proposed model is found to be adequate in order to represent a degradation and maintenance program. However it relies upon some strong assumptions, that might not be easy to verify or relate to when working with a real world problem. On the other have the numerical experiments showed that we can find reasonable results from simulations. Therefore the framework in this text might be used to explore what different changes in a maintenance program might lead to, through simulations.

It is also a kind of model that might become very appropriate as utilization of the continuously larger amount of the data capturing, becomes more apparent in reliability analysis. There are however a lot of work to be done in order in order to connect available measurements to statistical models. Further development of models like this one, must of course happen in cooperation between engineers and scientists from many different fields, like statistics, reliability and computer science. Many of the discussed suggestions for further work will also apply to similar models.

References

[1] Applebaum, David (2004). *Lévy Processes - From Probability to Finance and Quantum Groups,* Notices of the American Mathematical Society, Vol. 51, No. 11.

[2] Asmussen, Søren (2003). *Applied Probability and Queues. 2nd ed.* Springer Verlag New York

[3] Barbu, Stefan Vlad, Limnios Nikolaos, (2008). *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications*, Springer Science+Business Media, LLC.

[4] Casella, George and Berger, Roger L (2002). *Statistical Inference, second edition*, Brooks/Cole, Cengage Learning

[5] Castanier, B., Grall, A. and Bérenguer, C. (2005). *A condition-based maintenance policy with non-periodic inspections for a two-unit series system.* Reliability Engineering and System Safety 87 (2005) 109–120

[6] Chikkara, R. S., and Folks, J.L.(1989). *The Inverse Gaussian Distribution,* New York: Marcel Dekker.

[7] Cocozza-Thivent, Christiane (1997) *Processus stochastiques et fiabilité des systèmes.* Springer-Verlag Berlin Heidelberg

[8] Cox, D. R., Miller, H. D. (1965). *The Theory of Stochastic Processes*, Methuen Co LTD.

[9] Dieulle, L., Bérenguer, C., Grall, A., Roussignol M. (2003). *Sequential condition-based maintenance scheduling for a deteriorating system.* European Journal of Operational Research 150 (2003) 451–461

[10] Gilks, W.R., Richardson S. and Spiegelhalter D.J. (1996) *Markov Chain Monte Carlo in Practice,* Chapman & Hall

[11] Grall, Antoine, Dieulle Laurence, Bérenguer Cristophe and Roussignol, Michel (2002). *Continuous-Time Predictive-Maintenance Scheduling for a Deteriorating System.* IEEE Transactions on Reliability, Vol 51, No. 2, June 2002

[12] Lawless, Jerry and Crowder, Martin (2004), *Covariates and Random Effects in a Gamma Process Model with Application to Degradation and Failure.* Lifetime Data Analysis, 10, 213–227, 2004

[13] Lazar, Nicole (2013). *The Big Picture: Big Data Computing*, CHANCE, 26:2, 28-32.

[14] Lindqvist, Bo Henry and Skogsrud, Guro (2007). *Modeling of dependent competing risks by first passage times of Wiener processes.* IIE Transactions, 41:1, 72-80

[15] Løvoll, Grunde and Kadal, Jørgen Christian. (2014) *Big data - the new data reality and industry impact.* URL: https://www.dnvgl.com/technology-innovation/latest-publications.html

[16] Meeker, William Q., Hong, Yili (2013). *Reliability meets Big Data: Opportunities and Challenges*, Quality Engineering 26:1, 102-116.

[17] Michael, John R., Schucany William R and Haas, Roy W. (1976). *Generating Random Variates Using Transformations with Multiple Roots,* The American Statistician, Vol. 30, No. 2 (May, 1976), pp. 88-90

[18] van Noortwijk, J.M. (2007). *A survey of the application of gamma processes in maintenance.* Reliability Engineering and System Safety 94 (2009) 2–21.

[19] Rausand, Marvin and Høyland Arnljot (2004) *System Reliability Theory. Models, Statistical Methods and Applications. 2nd ed.* John Wiley & Sons, Inc.

[20] Ribeiro, Claudia and Webber Nick (2002). *Valuing Path Dependent Options in the Variance Gamma Model by Monte Carlo with a Gamma Bridge.* Journal Of Computational Finance

[21] Ribeiro, Claudia and Webber Nick (2003). *A Monte Carlo Method for the Normal Inverse Gaussian Option Valuation Model using an Inverse Gaussian Bridge.* Working paper

[22] Ross, Sheldon M. (2014). *Introduction to probability models, 11th edition,* Academic Press.

[23] Seshadri, V. (1993). *The Inverse Gaussian distribution,* Oxford science publications. Oxford University Press.

[24] Singpurwalla, Nozer D. (1995). *Survival in Dynamic Environments.* Statistical Science, Vol. 10, No. 1 (Feb., 1995), pp. 86-103

[25] Xiao Wang & Dihua Xu (2010). *An Inverse Gaussian Process Model for Degradation Data*, Technometrics, 52:2, 188-197, DOI:10.1198/TECH.2009.08197

[26] Zhi-Sheng Ye & Nan Chen (2014). *The Inverse Gaussian Process as a Degradation Model*, Technometrics, 56:3, 302-311, DOI:10.1080/00401706.2013.830074