Imperfect Maintenance Models : Steady-state, Heterogeneity and Applications

Xingheng Liu

Thèse
de doctorat
de l'UTT

Xingheng LIU

Imperfect Maintenance Models: Steady-state, Heterogeneity and Applications

Champ disciplinaire :
Sciences pour l'Ingénieur

2020TROY0011 Année 2020

Thèse en cotutelle avec Norwegian University of Science and Technology - Trondheim - Norvège

NTNU
Norwegian University of Science and Technology

utt
UNIVERSITÉ DE TECHNOLOGIE
TROYES

THESE

pour l'obtention du grade de

DOCTEUR
de l'UNIVERSITE DE TECHNOLOGIE DE TROYES

EN SCIENCES POUR L'INGENIEUR

Spécialité : OPTIMISATION ET SURETE DES SYSTEMES

présentée et soutenue par

Xingheng LIU

le 28 septembre 2020

Imperfect Maintenance Models: Steady-state, Heterogeneity and Applications

JURY

Mme Mitra FOULADIRAD	PROFESSEURE DES UNIVERSITES	Présidente
M. Christophe BÉRENGUER	PROFESSEUR DES UNIVERSITES	Rapporteur
M. Shaomin WU	FULL PROFESSOR	Rapporteur
M. Pierre-Etienne LABEAU	PROFESSEUR	Examinateur
M. Bo Henry LINDQVIST	FULL PROFESSOR	Examinateur
M. Thomas Michael WELTE	DOCTOR	Examinateur

Personnalités invitées

M. Yann DIJOUX	PROFESSEUR ASSOCIE UTT	Directeur de thèse
M. Jørn VATN	PROFESSOR	Directeur de thèse

Acknowledgements

I would like to thank my supervisors, Yann Dijoux and Jørn Vatn, for their valuable guidance during my thesis. From the academic perspective, they have always been patient and supportive whenever I need their help, providing me with constructive suggestions and opportunities for industrial collaborations. They gave me great freedom in the research topics and encouraged me to further explore diverse research directions. Without their encouragement, I shall never set foot in the academic filed.

I want to express my gratitude to the secretariat of the Doctor School of UTT, Isabelle Leclercq and Pascale Denis, and to Runa Nilssen, senior executive officer at the Faculty of Engineering of NTNU. They have made great effort to establish the cotutelle agreement between UTT and NTNU, and have ensured that everything went well regarding the courses and credits.

I greatly appreciate the administrative secretariat of UTT, André Bernadette and Veronique Banse, and of NTNU, Kari Elise Dhale. They have guaranteed me an enjoyable working environment.

I am very honored that Shaomin Wu and Christophe Bérenguer have accepted the hard task of rapporteur. I want to sincerely thank them and the jury members for the time they have spent to my work as well as for their valuable and enriching comments for the future.

A special thank to Maxim Finkelstein, professor at the University of the Free State, who has played an essential role in my first published paper and has given me numerous stimulating ideas about my research. His stringency and assiduousness inspire me, reminding me what it takes to become a researcher.

I would like to thank Håkon Toftaker, our collaborator at Bane NOR, who has answered all my questions about Norwegian railway, giving me access to the data I need and helping me establishing realistic models.

I am so grateful to my friends at UTT and at NTNU, particularly Jinrui, Nan, Xinge, Lin, Aibo, Liu, and Zehui, for not only the stimulating academic discussions and close cooperation, but also for the support and companion. The time we spent together is priceless, and the memories are just too sparkling to be forgotten.

Finally, I would like to thank my family, and particularly my mother, whom this work is dedicated to. No words can express my love and miss for you.

Xingheng Liu

Abstract

Over the years, maintenance has been gaining attention in so many different fields as people have begun to realize the importance of keeping systems and facilities working in a reliable state. From the production line to fighter aircraft, the cost related to maintenance activities—or the lack of it—is enormous. Systems' failures can result in as small as the loss of production or as large as a major accident. Consequently, instead of focusing on reactive maintenance, which aims to restore a system after an unexpected failure, people are more interested in preventive and predictive maintenances that could anticipate failures or catastrophic scenarios. However, as the systems are getting increasingly complex, scheduling an appropriate maintenance plan is not easy. The question is, what are the appropriate maintenance actions, and when to implement them?

To answer this question, one needs to have a profound understanding of the system's failure mechanism. A typical starting point is to find out if the system is deteriorating over time, and if yes, how to describe it? Could the failure occur randomly without any measurable forewarning, or are there potential signs indicating that failure is approaching? It is also crucial to investigate the effectiveness of maintenance activities that do not necessarily restore the system to the as-good-as-new state. In the worst-case scenario, maintenance could be harmful to the maintained device if it is inappropriately done. When formulating the maintenance strategy, both the aging of systems and the maintenance effectiveness should be considered.

This being, numerous mathematical models have been proposed to depict the failure/repair process that undergoes imperfect repairs. In the time dimension, the instant of failure/repair is a point on the real line; in the state dimension, the deterioration and restoration of the system are often modeled by the state transition, degradation path, or simply the variation of system's age. Naturally, the stochastic process is the central tool for such modeling.

The thesis is, therefore, prepared in this context. On the one hand, we investigate the application field of imperfect repair models while examining the consequences of inappropriate model fitting. On the other hand, new models, which we believe are more realistic, are established. The mathematical properties, e.g., distributions, correlations, asymptotic behaviors, are thoroughly studied. Particularly, the statistical inference is addressed as it is always essential to estimate the model parameters when fitting a model to the observations. The theoretical developments are accompanied by several case studies, based on not only simulated data but also data collected from the Norwegian railway network. We intend to highlight the relevance of using imperfect repair models in evaluating the reliability of systems, assessing the effectiveness of maintenance actions, and scheduling optimal maintenance plans. We hope that this work would be helpful and inspiring to practitioners and researchers.

Key words: Maintenance, Reliability, Service life (Engineering), Stochastic processes, Parameter estimation.

Résumé

Au fil des ans, la maintenance a attiré l'attention dans tant de domaines différents que les propriétaires et opérateurs commencent à réaliser l'importance de maintenir les systèmes et les installations dans un état de fonctionnement fiable. De la chaîne de production aux avions de chasse, les coûts liés aux activités de maintenance — ou leur absence — sont énormes. Les défaillances du système peuvent entraîner tout aussi bien des pertes mineures en production que des accidents majeurs. Par conséquent, au lieu de se concentrer sur la maintenance réactive qui vise à restaurer un système après une panne inattendue, les opérateurs sont de plus en plus intéressés par la mise en place de maintenances préventives et prédictives, qui peuvent anticiper et éviter les pannes ainsi que tout scénario catastrophique. Cependant, comme les systèmes deviennent de plus en plus complexes, établir un plan de maintenance approprié n'est pas une tâche facile. La question est, quelles sont les actions de maintenance pertinentes et à quel moment les mettre en œuvre?

Pour répondre à cette question, il faut avoir une compréhension approfondie du mécanisme de défaillance du système. Un point de départ usuel consiste à savoir si le système se détériore au fil du temps et si oui, parvenir à décrire le vieillissement. La panne se produit-elle de manière aléatoire et sans avertissement préalable et mesurable, ou y a-t-il des indicateurs permettant de détecter la défaillance? Il est également essentiel d'étudier l'efficacité des activités de maintenance qui ne restaurent pas nécessairement le système dans un état aussi bon que neuf. Dans le pire des cas, une maintenance peut être nuisible à l'appareil entretenu si elle n'est pas effectuée correctement. Lors de la formulation de la stratégie de maintenance, le vieillissement des systèmes et l'efficacité de la maintenance doivent être pris en compte.

Cela étant, de nombreux modèles mathématiques ont été proposés pour décrire le processus de défaillance/réparation d'un système qui subit des réparations imparfaites. En dimension temporelle, l'instant de défaillance/réparation est un point sur la ligne réelle; dans la dimension d'état, la détérioration et la restauration du système sont souvent modélisées par la transition d'état, le chemin de dégradation ou simplement la variation de l'âge du système. Naturellement, les processus stochastiques sont essentiels à cette modélisation.

La thèse est donc préparée dans ce contexte. D'une part, nous étudions le domaine d'application des modèles de maintenance imparfaite tout en examinant les conséquences d'un ajustement de modèle inapproprié. D'autre part, de nouveaux modèles, que nous pensons plus réalistes, sont mis en place. Les propriétés mathématiques, par exemple, les distributions, les corrélations, les comportements asymptotiques, sont soigneusement étudiées. En particulier, l'inférence statistique est abordée — lors de l'ajustement d'un modèle aux observations, il est toujours essentiel d'estimer les paramètres du modèle. Les développements théoriques sont accompagnés de plusieurs études de cas, basées non seulement sur des données simulées mais également sur des données collectées auprès du réseau ferroviaire norvégien. Nous avons l'intention de souligner la valeur des modèles de réparation imparfaits dans l'évaluation de la fiabilité des systèmes, l'évaluation de l'efficacité des réparations et l'élaboration de plans de maintenance, et nous espérons que ce travail sera utile et une source d'inspiration pour les praticiens et les chercheurs concernés.

Mots-clés : Entretien, Fiabilité, Durée de vie (ingénierie), Processus stochastiques, Estimation de paramètres.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Nowadays, many systems, such as production lines, weapon equipments, nuclear power stations, vehicles, aircrafts, etc., have become more and more complex. The costs of their usage are also getting even higher than before. Maintenance has to be carried out in order to keep these systems' performance close to the level of their original design. Most systems used in practice are subject to deterioration and aging with usage. For those wearing-out systems, maintenance, such as monitoring, repairs, and replacements, can extend their usage lifetimes, keep the quality of operations, reduce the cost of operations, and prevent system failures. Note that, in this work, we are only interested in repairable systems that can be restored to the operating state through component replacement or repair when a system failure occurs.

In practice, maintenance is divided into two major classes: one is corrective maintenance (CM), the other preventive maintenance (PM). Corrective maintenance aims to restore the system to a specified condition when the system fails. Preventive maintenance aims to retain the system in a specified condition or to improve the reliability of the system. With the ongoing trend of Industry 4.0, more and more effort is put into preventive, predictive, or even proactive maintenances. The goal is to minimize the potential loss of a system failure by carrying out preventive maintenance at the right time. Not only should we gather all kinds of data on the system's state, e.g., age and physical characteristics, which help us evaluate the remaining lifetime and the instantaneous failure intensity, but also a proper maintenance model is required to represent the interaction between maintenance activities and the physical system correctly.

Maintenance models are the basis of any quantitative maintenance analysis, which can be used to analyze and evaluate the performances of maintenance approaches. Several factors distinguish these maintenance models, including but not limited to, maintenance policies (age replacement, block replacement, failure limit...), system structures (series structure, parallel structure, k-out-of-n...), maintenance degree (perfect, minimal, imperfect...), optimization criteria (minimal cost rate, maximal availability, maximal reliability...), lifetime distributions (Exponential, Weibull, Gamma...).

The current dissertation focuses on imperfect maintenance models (IMMs) or imperfect repair models, wherein maintenance activities do not have to be perfect (restore the system to a good-as-new state) or minimal (bring the system back into the state just before the failure without any kinds of improvement). Although the word "imperfect" usually indicates that a system is brought back to a state between good-as-new and bad-as-old, these models are also capable of modeling harmful maintenance (worse-than-old) or over-perfect maintenance (better-than-new). Over the last decades, IMMs have been widely studied as they are flexible enough to represent a large scale of situations, which in turn leads to maintenance strategies that are more effective for minimizing the cost rate or maximizing the reliability/availability of assets. Here we list some of the main factors that characterize imperfect repair models:

- Type of maintenance. Both CM and PM can be imperfect. IMMs can involve only CM or both CM and PM.
- Maintenance effect. Maintenance is usually supposed to reduce the age of an item, but sometimes, they are assumed to reduce the failure intensity, or both.
- Stability. Some models are used to depict the deteriorating system with failures occurring more and more often, while others are utilized to model systems having a stationary state.
- Independence. In some models, the consecutive inter failure times are assumed to be independent, while in other models like the Brown-Proschan process, the inter failure times are correlated.
- Homogeneity of repairs. The imperfect repair is often characterized by a repair degree, which determines to what extent a system is restored. It could be a constant or evolve through time or be random.

IMMs are mathematically characterized by stochastic processes: this is not a surprise as an early (and still very popular) IMM, proposed by Kijima in the 1980s, bears the name of the "generalized renewal process." In the literature, many mathematical properties of IMMs have been exhibited: we can prove that certain models are stable while others are not; we can estimate the model parameters given data on the failure/repair process; there are statistical tests telling us whether a system may or may not be described by a certain model given the observed inter failure times. Meanwhile, numerous papers study the application of these models, e.g., finding the optimal PM policy when neither CM nor PM is perfect. However, some topics are not addressed, and some seemingly elementary questions should be answered.

1.2 Objectives

From a mathematical point of view, can we find out more about the IMMs? For example, in a renewal process, the distribution of the forward/backward recurrence time has long been known. Could it be generalized to stable IMMs? Moreover, if yes, how does it influence the maintenance policies?

IMMs are often applied to a single unit system or a homogeneous population composed of independent, identical systems. If the heterogeneity is present in the population, which is to say, the systems are similar but not identical, can we still estimate the model parameters correctly from the observations? How do we account for the heterogeneity, and how does it influence PM strategies?

Series systems are commonly used in reliability engineering. The system fails when one of its components stops functioning. If the failed component is replaced while others are not maintained, part of the system can be considered "renewed", making the replacement an imperfect repair at a system level. Can we use some simple IMMs to approximate series systems? And what are the potential gain and losses of such modeling compared to other approximation methods??

From a practical point of view, a maintenance manager usually needs to choose a proper maintenance model from several candidates. All of them are realistic but may provide different results when predicting the remaining lifetime, and may suggest different preventive maintenance strategies. How to identify the best (or the worst) IMM? How do we evaluate their performances?

This thesis aims further to discover IMMs both in theory and in practice. To be more specific, the main objectives include:

- Investigate the mathematical properties of stable IMMs.
- Model the heterogeneity and assess its influence on maintenance policies.
- Propose new models to approximate series systems.
- Evaluate and compare the performances of proposed models.

1.3 Research methods

The research is both theoretical and applied, exploratory and explanatory. New models are developed by combining knowledge from different fields, e.g., survival analysis and frailty analysis because the existing theories are not fully capable of explaining the observations. These models are then tested on field data, aiming to illustrate the potential benefits of their usage.

Mathematical deduction

The existing mathematical and statistical properties of IMMs, e.g., distributions, correlations, convergences, etc. form the foundation of our new theorems and propositions: all of them are based on strict mathematical proofs.

Nevertheless, some quantities, e.g., the asymptotic distribution of the inter failure time of a specific IMM or the long-run cost rate of repair activities when we want to evaluate and compare some PM strategies, can hardly be expressed by explicit formulas. In those cases, we resort to Monte Carlo simulation, as elaborated below.

Data collection and analysis

Only quantitative analysis is used during this work. Simulated data on the failure/repair process of a system is gathered via Monte Carlo simulation, with a specified model and known parameters. These data include typically the inter failure times (a positive number), lifetime type (censored or uncensored), repair type (CM or PM), repair degree (a proportion), etc. of a system described by a certain IMM. They have been utilized for the following purposes:

1. Verify the mathematical propositions and theorems. When a formula is proposed, it is first checked with simulated data, which allows a quick elimination of wrong guesses and formulas. Once the proof is complete, Monte Carlo simulation is used again to examine the validity of the propositions.

2. Verify the estimation procedures and evaluate the consistency and efficiency of the estimators. The statistical inference is one of the main issues of our research: when a model is fit to the observations, the underlying parameters have to be deduced by a certain estimation procedure, e.g., moment-matching, Maximum Likelihood Estimation (MLE). The validity of the inference procedures is examined by comparing the estimates to the known parameters. Furthermore, tuning the data size helps to evaluate the consistency and efficiency of the estimators.

3. Evaluate quantitatively maintenance strategies. We chose the long-run repair cost rate as the optimization criteria, which often involves the renewal function, i.e., the expected number of failures within a period of time. This is a typical quantity that does not possess an explicit formula. Simulating the failure/repair process of a large number of independent systems and recording the repair costs enables us to obtain the long-run repair cost rate empirically.

Field data from Bane NOR has also been analyzed. The failure/repair records of the Norwegian railway signaling system during the last decade contain information such as intervention date, repair duration, repair type, as well as the location of the signals, the manufacturer, and the administration. The data is then used to test and evaluate the existing IMMs as well as the proposed new models.

1.4 Research scope, limitations and future perspectives

The main focus of this work is on IMMs and quantitative maintenance optimization. Thus, many issues and modeling techniques regarding the reliability and risk analysis are not pursued. Below we list some of the main limitations within the framework of imperfect repair model:

- A binary representation of the system is used: they can be either working or failed. Therefore, multi-state modeling and the relevant approaches, e.g., Markov Chains, are not pursued. Under such a binary assumption, the models investigated in this

paper cannot be applied to systems that undergo continuous degradation, such as the corrosion process of pipes, the crack length of pavement, the geometrical deviation of railway tracks.

- The repair time is considered negligible. This is a quintessential assumption in reliability engineering, aiming to simplify the model when the repair time is negligible compared to its mean inter failure time. Although this assumption is valid for the data provided by Bane NOR in the sense that railway signals usually survive years, or even decades, while the repair activity generally lasts no more than a day, it limits the range of applications of our findings.

From a maintenance optimization point of view, this study has the following limitations:

- The maintenance optimization that we address in this study is merely one of the many steps in real maintenance management. It is a decision problem in the "ideal" situation. Application of such an "optimal" maintenance plans is generally not easy in practice. A typical problem could be a high backlog, which hinders the implementation of preventive maintenances.
- Covariates are not considered in this study. That is to say, all the reliability indicators of an asset, e.g., mean lifetime, failure frequency, aging speed, are derived uniquely from the lifetime data without using covariates such as meteorological data, train frequency, locations, etc..
- We deploy a purely quantitative method, which relies strongly on the quality and size of collected data. Basically, we derive the aging speed of the assets from the failure/repair records, and based on the estimated model parameters, a PM strategy is proposed. Therefore, statistical power when the data size is small should be questioned. Besides, a Bayesian framework that considers the expert's opinions is not pursued, but will be studied in the future.

To sum up, the limitations mentioned above point out some future research directions: combine the IMMs with multi-state or continuous state degrading system, where the repair duration or the stay time in different states is no longer negligible; incorporate covariates into the models; develop possibly a Bayesian framework to consider experts' opinion.

1.5 Dissertation structure

The rest of the dissertation comprises four chapters. In Chapter 2, we present state of the art on IMMs. Some of the most popular models are introduced in a detailed manner, along with proper mathematical definitions. Chapter 3 is dedicated to stable IMMs. We prove that they have similar mathematical properties as the renewal process does, before revealing the asymptotic distributions of the most crucial quantities, including virtual age, in IMMs. In Chapter 4, we discuss the heterogeneity in IMMs. The object under investigation is no longer a single system, but a group of similar systems that

undergo imperfect repair. Chapter 5 is dedicated to series systems, modeled by the superposition of renewal processes, and their approximations. We investigate the inner relationship between the series system and IMMs and propose original approximation approaches that help predict the remaining lifetime of series systems. In Chapter 6, the investigated models are tested on simulated data and field data. Notably, we address the maintenance optimization issue using the findings of Chapter 3 and 4. Concluding remarks and discussions are given in Chapter 7.

Some chapters in the main part of the paper have been peer reviewed. One can refer to [82] for Chapter 3, [83] for Sections 4.1 and 6.3, and [81] for Chapter 5 and Section 6.2.

Chapter 2

State of the art

Industrial systems are subject to repair actions when failures occur. As we adopt a binary representation of the system's state in this work, the gradual deterioration is not considered: a system can be either working or failed. Therefore, failures that result from external shocks or intrinsic aging are considered instantaneous events. To bring a damaged system back to work, repairs are carried out with a certain duration, depending on the complexity of the job. Nevertheless, the restoration (the state transition from failure to working) is also assumed to be instantaneous. Mathematically, failures and restorations are random events in time that could be represented by a collection of points located on some underlying mathematical space, e.g., a real line, and point processes are used to model these phenomena.

It has long been known that perfect repair and minimal repair correspond, in the framework of the point process, to renewal processes (RP) and non-homogeneous Poisson process (NHPP), respectively. Imperfect repairs, as elaborated in the following sections, are often modeled by generalizations of RP, including the virtual age process. In the rest of this chapter, the state of the art of IMMs are presented in the framework of point process: starting with the most basic notions, we will cover the topics of the renewal process, NHPP, virtual age process, superposition of renewal processes (SRP), geometric process, before addressing the issue of heterogeneity in survival analysis.

2.1 Point process and repair

The instantaneous events occurring randomly in time can be described by a point process $\{N_t, t \geq 0\}$ with a state space $\{0, 1, 2, ...\}$, where N_t represents the total number of events in $(0, t]$. The number of points located in $(s, t]$ with $s < t$ and $s, t > 0$ is $N_t - N_s$. Note that $N_s \leq N_t$ for $s \leq t$ and $N_0 = 0$.

A process has *independent increments* if the number of events (points) in disjoint intervals are independent random variables. A process has *stationary increments* if

$$N_t - N_s \quad \text{and} \quad N_{t+\tau} - N_{s+\tau}, \forall \tau > 0, \tag{2.1}$$

have the same distribution, and is therefore called a *stationary point process*. Throughout this study, we are interested only in *regular* process: there are no multiple occurrence or simultaneous observations.

A point process is fully defined by its intensity:

$$\forall t \geq 0, \lambda_t = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(N_{t+\Delta t} - N_{t^-} = 1 | \mathcal{H}_{t^-}), \tag{2.2}$$

where $\mathcal{H}_{t^-}$ is the corresponding natural filtration that records its "past behavior", i.e., the set of all points events in $(0, t]$.

Denote by $\{T_i\}, i \in \mathbb{N}$ the successive event times with $T_0 = 0$ and $\{X_i\}$ the inter-arrival times (also called intervals or cycles) with $X_0 = 0$ and $X_i = T_i - T_{i-1}$ for $i \geq 1$. The cumulative number of events up to t is denoted by

$$N_t = \sum_{i=1}^{\infty} \mathbb{1}_{\{T_i \leq t\}}, \tag{2.3}$$

where $\mathbb{1}_{\{T_i \leq t\}}$ is the indicator function that equals 1 if $T_i \leq t$, and is equal to 0 otherwise. Thus, the event occurrences over time can either be defined by $\{N_t\}$, a counting process that gives the number of incidents at any instant, or by $\{T_n, n = 0, 1, 2, ...\}$ the arrival points over $[0, \infty)$. In the following, the terms "point process" and "counting process" are used interchangeably.

2.1.1 Renewal process and perfect repair

The renewal theory [38] is widely applied in industry. For instance, the renewal function (that will be later defined) in practice can be interpreted as the mean number of replacements/perfect repairs for a system operating in a given interval of time. Thus, the mean number of the required spare parts can be estimated, so can be the probability of the spare parts shortage. Furthermore, when describing the performance of, for example, repairable production systems, renewal processes are generalized to the corresponding renewal reward processes that allow, among other things, to obtain the optimal, long-run maintenance policies [22].

In the rest of this section, some of the basic notions regarding the renewal process are introduced. Nevertheless, subjects that are not relevant to our contributions, e.g., delayed renewal process or equilibrium renewal process, are not addressed. Besides, the proofs of the theorems and propositions that appeared in this chapter are omitted, as they are mainly classic results in renewal theory. One can refer to [38] for the proofs.

The definition of a renewal process is presented below.

Definition 2.1 *A counting process $\{N_t, t \geq 0\}$ is called renewal process if its inter-arrival times, $\{X_i\}, i = 1, 2, ...$ are independent, identically distributed random variables.*

A perfect repair, which generally consists of replacing the failed component by a new identical one or bringing back the system to a good-as-new state, is, therefore, a renewal,

and the instants of perfect repairs coincide with the arrival times in the point process. The system is assumed to be as new at the time origin. Upon failure, it is instantaneously replaced by a new and identical one. The sequential inter-failure times are, therefore, i.i.d random variables, forming a renewal process.

Let $F(t)$ be the cumulative distribution function (Cdf) of a generic interval X, and $f(t)$ and $\lambda(t)$ be respectively the corresponding probability density function (Pdf) and failure rate. For a renewal process, the stochastic intensity defined by Eq.(2.2) reduces to:

$$\lambda_t = \lambda(t - T_{N_{t^-}}), \quad t \geq 0. \tag{2.4}$$

N_{t^-} is the number of events before instant t, $T_{N_{t^-}}$ is the time of the last renewal. $t - T_{N_{t^-}}$, also known as the backward recurrence time, is the time elapsed since the last renewal. It can be interpreted as the "age" of the system at time t. Similarly, $T_{N_{t^-}+1} - t$, also known as the forward recurrence time, is the time till next renewal and stands usually for the "remaining lifetime". Denote by $\overline{F}(t) = 1 - F(t)$ the survival function. Given the current system age x, the Cdf of the remaining lifetime δ of a system is

$$P(\delta \leq t|x) = F(t|x) = 1 - \overline{F}(t+x)/\overline{F}(x). \tag{2.5}$$

The asymptotic distributions of the forward and backward recurrence time, when t tends to infinity, will be further addressed.

Following the relationship between a counting process and its arrival times, we have

$$\{N_t < n\} \equiv \{T_n > t\}, \quad \{N_t = n\} \equiv \{T_n \leq t < T_{n+1}\}. \tag{2.6}$$

The probability of having exactly n events in interval $(0, t]$ is, therefore,

$$\begin{aligned} P(N_t = n) &= P(N_t \geq n) - P(N_t \geq n+1) \\ &= P(T_n \leq t) - P(T_{n+1} \leq t) \\ &= F_n(t) - F_{n+1}(t), \end{aligned} \tag{2.7}$$

where $F_n(t)$ is the n-fold convolution of $F(t)$ with itself.

Renewal function

Definition 2.2 *The renewal function is defined by the expectation:*

$$M(t) = E[N_t]. \tag{2.8}$$

This is the mean number of events within the period $(0, t]$. It can also be expressed by the sum of convolutions:

$$M(t) = \sum_{n=1}^{\infty} F_n(t). \tag{2.9}$$

Let $m(t) = M'(t)$. $m(t)$ is called the *renewal density function*:

$$m(t) = \sum_{n=1}^{\infty} f_n(t), \tag{2.10}$$

where $f_n(t) = dF_n(t)/dt$. $m(t)dt$ could be interpreted as the probability of a renewal occurring in $(t, t + dt]$.

Limiting properties

Let μ be a finite mean inter-arrival time with $\mu \equiv E[X] = \int_0^\infty \overline{F}(u)du < \infty$. For convenience purpose, throughout the chapter, we assume that X is continuous. The following theorem holds [95]:

Theorem 2.1 *When* $t \to \infty$,

$$\frac{N_t}{t} \to \frac{1}{\mu}, \text{ a.s.}, \tag{2.11}$$

and

$$\frac{M(t)}{t} \to \frac{1}{\mu}. \tag{2.12}$$

Relation (2.12), also known as *elementary renewal theorem*, has an intuitive interpretation: the mean cycle duration, μ, is approximately t over number of renewals when t tends to infinity, according to the strong law of large numbers.

Key renewal theorem

Theorem 2.2 *Let* $M(t)$ *be the renewal function with finite mean* μ. *For any directly Riemann integrable function* $h(s)$*:*

$$\lim_{t\to\infty} \int_0^t h(t-x)dM(x) = \frac{1}{\mu} \int_0^\infty h(s)ds. \tag{2.13}$$

The key renewal theorem is one of the possible approaches to derive the asymptotic distributions of forward/backward recurrence time, as elaborated below. It is also indispensable for the theoretical developments for stable imperfect repair processes. For more comprehensive discussions, one can refer to [95] and [24].

Renewal equation

Suppose that $a : [0, \infty) \to \mathbb{R}$ is locally bounded. Let $*$ be the convolution operator. An integral equation of the form

$$u = a + u * F, \tag{2.14}$$

for an unknown function $u : [0, \infty) \to \mathbb{R}$ is called a *renewal equation* for u. Denote by F the Cdf of the inter-arrival times X. The renewal equation for the renewal function M is given by:

$$M = F + M * F, \tag{2.15}$$

or, more commonly,

$$M(t) = F(t) + \int_0^t M(t-x)dF(x). \tag{2.16}$$

Eq.(2.16) can be easily proved by conditioning on the time of the first renewal. If the first cycle $x \leq t$, then the process "restarts" and the expected number of renewals in the interval $(x, t]$ is $M(t-x)$. The solution of the renewal equation by applying the Laplace transform to Eq.(2.15). Yet, the explicit formula for $M(t)$ can only be derived for some specific cases, e.g., when the cycle X is gamma/exponential distributed, and often one needs to resort to numerical solutions or approximations. Some recent literatures on approximation of the renewal type equation includes [92], [90], [59] and [58].

Age, residual lifetime and spread

At a chronological time t, the backward recurrence time $B(t)$, also known as age, as well as the forward recurrence time $\delta(t)$, also known as residual lifetime, are given by:

$$B(t) = t - T_{N_t}, \tag{2.17}$$

$$\delta(t) = T_{N_t+1} - t. \tag{2.18}$$

The age and residual lifetime have the same limiting distribution [95] as t tends to infinity. Denote by μ the mean cycle duration and $F(t)$ the Cdf, then,

$$\lim_{t\to\infty} P(B(t) \leq y) = \lim_{t\to\infty} P(\delta(t) \leq y) = \frac{1}{\mu}\int_0^y \overline{F}(u)du. \tag{2.19}$$

Let $Y(t) = B(t) + \delta(t)$. $Y(t)$ represents the duration of the cycle that contains t, and is called *spread*, or total life. The distribution of $Y(t)$ as t tends to infinity is given by

$$\lim_{t\to\infty} P(Y(t) \leq y) = \frac{1}{\mu}\int_0^y sf(s)ds. \tag{2.20}$$

Obviously, Eq.(2.20) differs from $F(t)$, meaning that if we observe at instant t a renewal process that started a long time ago, then the observed cycle, i.e. the cycle containing t, does not share the same distribution as other cycles. In fact, as t tends to infinity, $Y(t)$ is stochastically larger than X. This is known as the *inspection paradox*. The definition of usual stochastic order [97] is given below.

Definition 2.3 *A random variable X is said to be stochastically less (or equal to) Y, written $X \leq_{ST} Y$, if the upper tail probability satisfies:*

$$P(X > t) \leq P(Y > t), \quad t \geq 0. \tag{2.21}$$

The inspection paradox has an intuitive explanation: parts with a longer actual life are more likely to be observed than parts with a shorter actual life. Specifically, in a Poisson process, the expected value of $Y(t)$ as t tends to infinity is double the average lifetime, due to the memorylessness of exponential distribution. Some extensions and implications of the inspection paradox are addressed in [68] and [115].

Alternating renewal process and reward renewal process

Ordinary renewal processes are defined assuming that the repair durations can be neglected. In practice, however, this is generally not true, although sometimes the lifetime of the system is considerably larger than the repair duration. Let $\{U_i\}, i = 1, 2, ...$ be the i.i.d. lifetimes of a system and $\{V_i\}, i = 1, 2, ...$ be i.i.d. repair durations. Then, $\{Z_i = U_i + V_i\}, i = 1, 2, ...$ is called *alternating renewal process*.

The *availability* $A(t)$ represents the probability that the system is operating at instant t. While in practice, the limiting availability $A = \lim_{t\to\infty} A(t)$, is more widely used since it shows the proportion of system being in the working state in the long term.

Let μ_u and μ_v be respectively the expected value of U and V. The limiting availability A of a system in accordance with the above described alternating renewal process is given by:

$$A = \frac{\mu_u}{\mu_u + \mu_v}. \tag{2.22}$$

The instant availability $A(t)$ is usually calculated via numerical approaches. However, for the simplest case where the lifetime U and repair time V are both exponentially distributed with failure rate $1/\mu_u$ and repair rate $1/\mu_v$, the explicit form exists [56]:

$$A(t) = \frac{\mu_u}{\mu_u + \mu_v} + \frac{\mu_v}{\mu_u + \mu_v} e^{-(1/\mu_u + 1/\mu_v)t}. \tag{2.23}$$

The notion of stopping time should be introduced before addressing the reward renewal process since it is indispensable for proving Wald's equality. Stopping time is usually defined by a *stopping rule*, a mechanism for deciding whether to continue or to stop the process based on the present position and past events. In other words, the stopping time is uniquely determined by history, as defined below.

Definition 2.4 *Let $X_n, n = 1, 2, ...$ be a sequence of random variables. An integer valued random variable N is called a stopping time for X_n if for all $n = 1, 2, ...$, event $\{N = n\}$ is independent of $X_{n+1}, X_{n+2}....$*

In this regard, the number of observed events N is a stopping time for a renewal process, and

Theorem 2.3 *(Wald's equality) If N is a stopping time with $E[N] < \infty$ for a renewal process with finite mean inter-arrival times,*

$$E\left[\sum_{i=1}^{N} X_i\right] = E[N]E[X]. \tag{2.24}$$

Maintenance cannot be carried out without needed resources: manpower, material, administration...Repairs are therefore always accompanied by a cost, which is referred to as, more generally, *reward*. Denote by R_n the reward after the n-th cycle. Assume that R_n are i.i.d. random variables with finite mean $E[R]$, and are independent from the cycles, X. The total reward in $(0, t]$ is therefore by the reward process:

$$R(t) = \sum_{n=1}^{N_t} R_n, \tag{2.25}$$

and

$$\lim_{t\to\infty} \frac{E[R(t)]}{t} = \frac{E[R]}{\mu}. \tag{2.26}$$

Eq.(2.26) shows that the long-run reward per unit of time, or long-run reward rate, is simply the mean reward over the mean cycle duration. We will present next some classic maintenance optimization problems wherein the reward is the repair cost.

Classic maintenance optimizations with renewal process

Since one of the major issues addressed in this work is maintenance optimization with imperfect repairs, it is crucial to discuss first some basic maintenance strategies with perfect repair, i.e., when the consecutive restorations form a renewal process. The basic assumptions on the failures and corrective maintenances (CM) include:

- Instantaneous failure: no continuous deteriorating or multiple state between working and failure.
- Instantaneous CM: no delay between the failure and the maintenance job.
- Maintenance duration is neglected.
- Perfect CM: systems restored to good-as-new state.
- Increasing failure intensity.

The first four assumptions guarantee that the CM forms a renewal process when no preventive maintenance (PM) is implemented, while the fifth assumption, increasing failure rate, eliminates the situations where the system is not aging as time goes by; otherwise, no PM is needed. Most industrial units and maintenance managers are looking for effective approaches to avoid the breakdown of a system because responding to an unexpected failure generally costs more than replacing *in advance* the unit that is likely to fail, with all the resources well prepared. In this context, hundreds of papers and books have been published on maintenance optimization. One can refer to [25], [96], [104] for some most recent review papers. Here, we specifically address two simple PM policies that have been studied by many other researchers under the framework of imperfect repair and will be thoroughly developed in Section 4 and 6 of this monograph: block replacement and age-based replacement policy.

Block replacement policy is a type of preventive replacement policy where the units are always replaced at failure or at a scheduled time periodically [65]. Dating back to 1960 [4], it is still one of the most popular and widely applied PM policy due to its ease of administering.

Let C_c be the cost of CM and C_p that of PM, with $C_c > C_p$. The PM is implemented at scheduled and fixed intervals: $\tau, 2\tau, 3\tau....$ We assume that the PM is perfect and that the costs of CM and PM are constant. In practice, this is not necessarily the case: they could under certain circumstances be varying with time, e.g., if the difficulty of maintenance is gradually increasing and more resources are needed. Let $M(t)$ be the corresponding renewal function when no PM is involved. Following Eq.(2.26), the long-run average cost per unit of time is given by:

$$c(\tau) = \frac{C_p + C_c \cdot M(\tau)}{\tau}. \tag{2.27}$$

Due to the intractability of $M(t)$, the approximate form is more commonly used: assuming that it is not probable to have more than one CM within a PM interval, i.e., $P(N_\tau > 1) = 0$, then, $M(t)$ in Eq.(2.27) could be replaced by $F(t)$.

Another typical PM is called age-based replacement policy. A system whose lifetime X is described by the Cdf $F(t)$, the pdf $f(t)$ and the failure rate $\lambda(t)$ is replaced at time T from the last renewal point or at failure, whichever comes first. The idea is, block replacement sometimes generates unnecessary waste: a PM could be nearly useless if a CM occurred just before. Therefore, it may be wise to wait till that the age of the unit reaches a predefined threshold, T, forcing the PM to keep its distance from the previous maintenances.

The duration of a renewal cycle is therefore $min(X, T)$, with the expectation $\int_0^T \overline{F}(u)du$. The expected cost on the renewal cycle is $C_p\overline{F}(T) + C_cF(T)$. Using the renewal reward Theorem 2.26, the long-run repair cost per unit of time is given by:

$$c(T) = \frac{C_p\overline{F}(T) + C_cF(T)}{\int_0^T \overline{F}(u)du}. \tag{2.28}$$

2.1.2 Minimal repair and non-homogeneous Poisson process

First introduced by [4] in 1960, the minimal repair was studied and applied in numerous fields, such as the modeling of repair and maintenance or bio-demographic studies [108]. In contrast to perfect repair, the minimal repair usually describes a minor maintenance or repair operation that does not make the system "younger". Let a system be described with Cdf $F(t)$ and failure intensity $\lambda(t)$. At age x, the system fails and is instantaneously minimally repaired if its survival function after the repair is [57]

$$\frac{\overline{F}(x+t)}{\overline{F}(x)} = e^{-\int_x^{x+t} \lambda(u)du}. \tag{2.29}$$

Eq. (2.29) indicates that after the repair, the failure intensity is just $\lambda(x)$, the same as it was prior to the maintenance. A minimal repair has, therefore, no impact on the

future stochastic behavior of the system, as if a failure did not occur. This is described as the repair that restores the system to a state it had been in prior to the failure; in other words, *as-bad-as-old*. A common interpretation of minimal repair is, when a complex system consisting of many independent parts, replacing one of them could be considered as minimal repair at a system level since the proportion of system being renewed is relatively small.

Now, let's consider it from the perspective of a stochastic process. First, we give the proper definition of a non-homogeneous Poisson process (NHPP).

Definition 2.5 *A counting process* $\{N_t, t \geq 0\}$ *is called a non-homogeneous Poisson process with intensity* $\lambda(t)$*, if*
(I) $N_0 = 0$*;*
(II) The process $\{N_t, t \geq 0\}$ *has independent increments;*
(III) $P(N_{t+h} - N_t = 1) = \lambda(t)h + o(h)$*;*
(IV) $P(N_{t+h} - N_t \geq 2) = o(h)$*.*

The stationary increments property as in homogeneous Poisson process is no longer conserved under the third assumption: the distribution of the i-th inter-arrival time depends on its predecessors, i.e., the sum of all the previous intervals. Hence, the cycles are not necessarily i.i.d random variables as in a renewal process. Using the properties (ii) and (iii) of definition 2.5, it is straightforward to prove that the stochastic intensity of an NHPP is:

$$\lambda_t = \lambda(t), \quad t \geq 0. \tag{2.30}$$

Denote by $\Lambda(t)$ the cumulative intensity with $\Lambda(t) = \int_0^t \lambda(u)du$, the number of events within the interval $(u, u+t]$ is distributed as:

$$P(N(u+t) - N(u) = n) = \frac{(\Lambda(u+t) - \lambda(u))^n}{n!} e^{-(\Lambda(u+t)-\lambda(u))}, \tag{2.31}$$

and the expected number of events within the interval $(0, t]$ is:

$$E[N(t)] = \Lambda(t). \tag{2.32}$$

In reliability applications, when a system having an increasing failure intensity undergoes minimal repairs, failures shall appear more and more frequently. Thus, an overhaul that rejuvenates the whole system completely is usually required, and PM policies should be made accordingly. Particularly, for NHPP, the previously defined block replacement coincide with the aged-based PM definition: replacing the system at age T is equivalent to a periodic replacement with period T, since the system's age is not influenced by CM, and is set to 0 at PM. The corresponding long-run average repair cost rate is given by:

$$c(\tau) = \frac{C_p + C_c \cdot \Lambda(\tau)}{\tau}. \tag{2.33}$$

2.1.3 Brown Proschan model

Having introduced the perfect and minimal repair, let us look into a basic imperfect repair model, the Brown Proschan model, or BP process. Proposed in [6] and [15], a BP process combines the perfect repair with minimal repair by assuming that, when a failure occurs, a repair is either perfect with probability p or minimal with probability $1-p$, and that the type of maintenance is independent from the past maintenances types. This model is relevant, e.g., in situations where some of the minor failures of a complex system are minimally repaired, whereas other, more serious failures result in an overhaul of the failed system. Theoretical results on the BP model and the corresponding statistical inference, as well as identifiability issues, have been addressed in [111], [55], [79], and [76]. Here, we focus particularly on the distributions and asymptotic properties.

The repair effects can be defined by i.i.d random variables $B_n, n \in \mathbb{N}^*$ that follows a Bernoulli distribution:

$$B_n = \begin{cases} 1, & \text{if the n-th repair is perfect;} \\ 0, & \text{if the n-th repair is minimal.} \end{cases} \tag{2.34}$$

Let $T_n, n \in \mathbb{N}$ be the occurrence times of a BP process and $X_n, n \in \mathbb{N}$ the inter-arrival times. The failure rate is $\lambda(t)$. Denote by $A_n, n \in \mathbb{N}$ the age just after the n-th repair with $A_0 = 0$. If the system is perfectly repaired, it is brought back to an as-good-as-new state, and the age is reduced to 0. Otherwise, it is minimally repaired, and A_n is the sum of X_n, the latest interval and A_{n-1}, the previous age:

$$A_n = (1 - B_n)(X_n + A_{n-1}). \tag{2.35}$$

Let $\sum_{j=k}^{l} \cdot = 0$ and $\prod_{k=j}^{l} \cdot = 1$ for $k > l$, then

$$A_n = \sum_{j=1}^{n} \left[\prod_{k=j}^{n} (1 - B_k) \right] X_j. \tag{2.36}$$

At time t, the system is equivalent to a new system that has been in operation for the time elapsed since the last perfect repair. Therefore, its stochastic intensity verifies:

$$\lambda_t = \lambda\left(t - T_{N_{t^-}} + \sum_{j=1}^{N_{t^-}} \left[\prod_{k=j}^{N_{t^-}} (1 - B_k) \right] X_j \right). \tag{2.37}$$

The survival function of a system with age x is showed in Eq.(2.5). The distribution of each interval in a BP process depends on the age just before the latest repair, which in turn depends on all the intervals after the latest perfect repair, as suggested by Eq.(2.35). The intervals X_n are, therefore, neither independent nor identically distributed. The distribution of the age A_n should be addressed before showing that of X_n.

Theorem 2.4 *[76] The Cdf of age A_n in a BP process characterized by p, the probability*

of perfect repair and $\Lambda(t)$*, the cumulative intensity function is given by:*

$$F_{A_n}(t) = 1 - (1-p)e^{-\Lambda(t)}\left[\sum_{k=0}^{n-1}(1-p)^k\frac{\Lambda^k(t)}{k!}\right]. \tag{2.38}$$

Consequently, the ages A_n converge in distribution: if $p > 0$, $A_n \xrightarrow{\mathscr{D}} A_\infty$, and

$$F_{A_\infty}(t) = 1 - (1-p)e^{-\Lambda(t)}. \tag{2.39}$$

Similar convergence results have been proved by Last and Szekli [75] for a larger class of imperfect repair models. The expected values of age are obtained by integrating the corresponding survival functions:

$$E[A_n] = (1-p)\int_0^\infty e^{-\Lambda(x)}\left[\sum_{k=0}^{n-1}(1-p)^k\frac{\Lambda^k(x)}{k!}\right]dx, \tag{2.40}$$

and

$$E[A_\infty] = (1-p)\int_0^\infty e^{-\Lambda(x)}dx. \tag{2.41}$$

The survival function of intervals X_n can be deduced from theorem 2.4. For $n \geq 1$,

$$R_{X_{n+1}}(t) = (1-p)^n\int_0^\infty \lambda(u)\frac{\Lambda^{n-1}(u)}{(n-1)!}e^{-\Lambda(t+u)}du + p\left[\sum_{i=0}^{n-1}(1-p)^i\int_0^\infty \lambda(u)\frac{\Lambda^{i-1}(u)}{(i-1)!}e^{-\Lambda(t+u)}du\right] + pe^{-\Lambda(t)}, \tag{2.42}$$

and, if $p > 0$, $X_n \xrightarrow{\mathscr{D}} X_\infty$, with

$$R_{X_\infty}(t) = p\int_0^\infty \lambda(t+u)e^{-\Lambda(t+u)+(1-p)\Lambda(u)}du. \tag{2.43}$$

The expected values are given by

$$E[X_n] = \int_0^\infty e^{-\Lambda(u)}\left[(1-p)^{n-1}\frac{\Lambda^{n-1}(u)}{(n-1)!} + p\sum_{k=0}^{n-2}(1-p)^k\frac{\Lambda^k(u)}{k!}\right]dx, \tag{2.44}$$

and

$$E[X_\infty] = p\int_0^\infty e^{-p\Lambda(x)}dx. \tag{2.45}$$

The perfect repairs in a BP process form an embedded renewal process [15]. Let $\lambda_p(t)$ and F_p be respectively the failure rate and Cdf of the duration between two perfect repairs. Then,

$$\lambda_p(t) = p\lambda(t), \tag{2.46}$$

and

$$\overline{F}_p(t) = \overline{F}^p(t). \tag{2.47}$$

These properties will be further used in Chapter 3. More industrial applications including specifically the maintenance optimization problems, could be found in, [70], [74], [30], [31] and [45], to name a few. Other extensions of the BP process were addressed in, among others, [9], where the probability of perfect repair is an age-dependent function $p(t)$, and in [7] where multiple failure types were considered.

2.1.4 Virtual age process

Among imperfect maintenance models, virtual age (VA) models [67] are some of the most popular ones, wherein the aging of a repairable system is assumed to depend on VA, which can have a value between zero and the operating time. We will start with definitions, interpretations, and meanings of virtual age before addressing the general repair process and their limiting properties.

Let a new system with a lifetime T be described by the Cdf $F(t)$ and failure rate $\lambda(t)$. It starts working at $t = 0$. Then, at the age of x, the Cdf of the remaining lifetime is given by $F(t|x) = 1 - \overline{F}(t + x)/\overline{F}(x)$. Assume that after the instantaneous maintenance (corrective or preventive) carried out at time t, the remaining lifetime is defined as the lifetime of a new and unmaintained system having age y. Then y is called the *virtual age*, the calendar age after this operation is, obviously, still t.

A repair with a negligible duration is carried out immediately after T_i and is supposed to reduce a system's age to $A_i, A_0 = 0$, which is called the virtual age or effective age after the i-th failure. Then, the remaining lifetime of the repaired system does not depend on the entire failure/repair history but depends on the virtual age of the system after the most recent repair. Mathematically, this is described by

$$P\left(X_{i+1} \leq t | T_1, T_2 \ldots T_i\right) = F(t|A_i), \quad \forall i \in \mathbb{N}, \quad \forall t \geq 0. \tag{2.48}$$

Thus, a virtual age process is fully defined by the age reduction mechanism, which determines the virtual ages $\{A_i\}$, and by $F(t)$, the Cdf of a new system. N_{t^-} being the number of failure before t, the virtual age at time t is given by

$$V_t = t - T_{N_{t^-}} + A_{N_{t^-}}, \tag{2.49}$$

and the stochastic intensity of the VA process is

$$\lambda_t = \lambda(t - T_{N_{t^-}} + A_{N_{t^-}}). \tag{2.50}$$

The failure rate of a new system, $\lambda(t)$, is particularly referred to in the context of the VA process as *baseline failure rate*. As the shape of the baseline failure rate is fixed, the virtual age at the start of a cycle is uniquely defined by the "position" of the corresponding point on the failure rate curve after the repair.

General repair

Let the age reduction mechanism of an imperfect repair be described by a function $q(x)$, where $q(x)$ is increasing, satisfying $0 \leq q(x) \leq x$. The harmful repair ($q(x) > x$, system "older" than before) and better-than-new repair ($q(x) < 0$) are not considered. The sequence of virtual ages $\{A_n\}$ at the start of the cycles are therefore given by:

$$A_n = q(A_{n-1}, X_n), \quad n \in \mathbb{N}^*. \tag{2.51}$$

The specific linear case where $q(x) = (1-\rho)x, 0 < \rho < 1$, is generally referred to as Kijima Type II model with constant repair degree [66], although it has early been considered descriptively in [15]. ρ represents the repair efficiency or restoration factor: the larger ρ, the more effective the repair is. Even though it seems very simple, the constant restoration factor is widely used in the literature and can describe the maintenance effect in many real industrial cases. In [86], the constant restoration factor is used to describe the effect of planned maintenance on systems' virtual age. Other papers considering the constant repair efficiency includes [51], [84], [117] and [50], to name a few. Kijima Type I models, on the other hand, assumes that the reduced amount of virtual age after a repair is proportional to the last inter-failure time [103] [5], i.e., $A_n = A_{n-1} + q(X_n)$. These two models were unified in [34] by introducing the model of *Arithmetic Reduction of Age with memory m*:

Definition 2.6 *The Arithmetic Reduction of Age with memory m (ARA_m) is defined by the stochastic intensity:*

$$\lambda_t = \lambda\left(t - \rho \sum_{j=0}^{min(m-1, N_{t^-}-1)} (1-\rho)^j T_{N_{t^-}-j}\right). \tag{2.52}$$

When $m = 1$, Eq.(2.52) reduces to

$$\lambda_t = \lambda(t - \rho T_{N_{t^-}}), \tag{2.53}$$

which indicates that the virtual age after the latest repair is $(1-\rho)T_{N_{t^-}}$. In other words, the VA after the n-th repair can be calculated iteratively by

$$A_n = A_{n-1} + (1-\rho)X_n. \tag{2.54}$$

Therefore, ARA_1 corresponds to Kijima Type I model with constant restoration factor. Consider now the case where $m = \infty$. The stochastic intensity in an ARA_∞ process is given by:

$$\lambda_t = \lambda\left(t - \rho \sum_{j=0}^{N_{t^-}-1} (1-\rho)^j T_{N_{t^-}-j}\right), \tag{2.55}$$

which indicates that the VA after the n-th repair can be expressed by

$$A_n = (1-\rho)^n A_0 + \sum_{j=1}^{n} (1-\rho)^{n-j+1} X_j, \tag{2.56}$$

or, in an iterative form,

$$A_n = (1 - \rho)(A_{n-1} + X_n). \tag{2.57}$$

Thus, ARA_∞ corresponds to Kijima Type II model with constant repair efficiency. These two models represent two marginal cases of history for the corresponding stochastic repair processes, i.e., the history that "remembers" all the previous repair times and that "remembers" only the last repair time. Maintenance scheduling and optimizations based on Kijima models have been investigated in [29], [60], [67], [84], and [85].

Aging and limiting properties

Last and Szekli [75] have proven the convergence of the Kijima Type II model, and hence of the ARA_∞ model to a steady-state regime under relative weak conditions. Finkelstein [39] has proven the convergence of the ARA_∞ model to a steady-state regime when the repair efficiency depends on the chronological age of the system. To begin with, consider the sequence formed by consecutive virtual ages.

Theorem 2.5 *[57] The virtual ages just after the n-th repair defined in Eq.(2.51),* $\{A_n\}$*, form a stochastically increasing sequence:*

$$A_n <_{ST} A_{n+1}, \quad n \in \mathbb{N}^*. \tag{2.58}$$

Under such an imperfect repair, the system seems to become "older and older" when more and more failures occur. However, the following theorem shows that the virtual ages converge and become stable, eventually.

Theorem 2.6 *[57] Assume that the baseline failure rate is increasing. Then, the virtual ages defined in Eq.(2.51),* $\{A_n\}$*, converge in distribution:*

$$A_n \xrightarrow{\mathscr{D}} A_\infty. \tag{2.59}$$

A direct consequence of this theorem is the convergence of the cycles $\{X_n\}$ [57], given that the distribution of X_n depends uniquely on the age after the latest repair, as shown in Eq.(2.48):

$$X_n \xrightarrow{\mathscr{D}} X_\infty. \tag{2.60}$$

In contrast, Doyen [33] has proven that the ARA_1 model behaves asymptotically as a non-homogeneous Poisson process. In fact, in ARA_1 model, the virtual age is strictly increasing. A major difference between the ARA_1 and ARA_∞ models could be therefore highlighted when the restoration factor is in the interval $(0, 1)$. If the failure rate of a new system is increasing monotonically to infinity, the inter-failure times converge to zero for the ARA_1 model and to a stationary distribution for the ARA_∞ model.

Weibull ARA_∞

Nguyen et al. [89] considered the special case at which an ARA_∞ obeyed a Weibull baseline distribution with shape parameter β and pseudo-scale parameter α:

$$F(t) = 1 - e^{-\alpha t^\beta}, \quad \alpha > 0, \beta > 0. \tag{2.61}$$

The corresponding baseline failure intensity, $\lambda(t)$, is given by

$$\lambda(t) = \alpha\beta t^{\beta-1}. \tag{2.62}$$

The Weibull distribution is widely applied in industry due to its flexibility to model different shapes of monotonic failure rate: $\lambda(t)$ is increasing when $\beta > 1$, decreasing when $\beta < 1$ and is a constant when $\beta = 1$. α is called pseudo-scale parameter because the true scale parameter is often denoted by η, with $\alpha = \eta^{-\beta}$. The Weibull ARA_∞ process is thus fully determined by the triple (α, β, ρ). Let $q = (1-\rho)^\beta$. The survival functions of X_n and A_n, denoted respectively by $R_{X_n}(t)$ and $R_{A_n}(t)$, are given by [89]

$$R_{A_n}(t) = \sum_{k=1}^{n} \frac{1}{(q,q)_{n-k}\left(\frac{1}{q},\frac{1}{q}\right)_{k-1}} e^{-\frac{\alpha t^\beta}{q^k}}, \quad n \in \mathbb{N}^*, \tag{2.63}$$

and

$$R_{X_{n+1}}(t) = \sum_{k=1}^{n} \frac{\int_0^\infty \alpha\beta x^{\beta-1} e^{-\alpha(x+t)^\beta + \alpha\left(1-q^{-k}\right)x^\beta} \mathrm{d}x}{q^k (q,q)_{n-k}\left(\frac{1}{q},\frac{1}{q}\right)_{k-1}}, \quad n \in \mathbb{N}^*. \tag{2.64}$$

where $(a,q)_k = \prod_{j=0}^{k-1}(1-aq^j)$ is the q-Pochhammer symbol. For $n = 0$, $R_{X_1}(t) = e^{-\alpha t^\beta}$. By integrating the survival functions over $[0,\infty)$ we can derive the expected values of the cycles and virtual ages:

$$E[A_n] = \alpha^{-\frac{1}{\beta}}\Gamma(\frac{1}{\beta}+1)\sum_{k=1}^{n} \frac{q^{\frac{k}{\beta}}}{(q,q)_{n-k}\left(\frac{1}{q},\frac{1}{q}\right)_{k-1}}, \quad n \in \mathbb{N}^*, \tag{2.65}$$

$$E[X_{n+1}] = \alpha^{-\frac{1}{\beta}}\Gamma(\frac{1}{\beta}+1)\sum_{k=1}^{n+1} \frac{q^{\frac{k-1}{\beta}}(1-q^{\frac{1}{\beta}}+q^{n+1-k+\frac{1}{\beta}})}{(q,q)_{n+1-k}\left(\frac{1}{q},\frac{1}{q}\right)_{k-1}}, \quad n \in \mathbb{N}. \tag{2.66}$$

The limiting distributions as n tends to infinity are denoted as R_{X_∞} and R_{A_∞} and can be expressed as follows:

$$R_{A_\infty}(t) = \sum_{k=1}^{\infty} \frac{1}{(q,q)_\infty\left(\frac{1}{q},\frac{1}{q}\right)_{k-1}} e^{-\frac{\alpha t^\beta}{q^k}}, \tag{2.67}$$

$$R_{X_\infty}(t) = \sum_{k=1}^{\infty} \frac{\int_0^\infty \alpha\beta x^{\beta-1} e^{-\alpha(x+t)^\beta + \alpha\left(1-q^{-k}\right)x^\beta} \mathrm{d}x}{q^k (q,q)_\infty\left(\frac{1}{q},\frac{1}{q}\right)_{k-1}}. \tag{2.68}$$

The expected values of X_∞ and A_∞ are also stated below.

$$E[A_\infty] = \alpha^{-\frac{1}{\beta}}\Gamma(\frac{1}{\beta}+1)\sum_{k=1}^{\infty}\frac{q^{\frac{k}{\beta}}}{(q,q)_\infty\left(\frac{1}{q},\frac{1}{q}\right)_{k-1}}, \tag{2.69}$$

$$E[X_\infty] = \alpha^{-\frac{1}{\beta}}\Gamma(\frac{1}{\beta}+1)\sum_{k=1}^{\infty}\frac{q^{\frac{k}{\beta}}(q^{-\frac{1}{\beta}}-1)}{(q,q)_\infty\left(\frac{1}{q},\frac{1}{q}\right)_{k-1}}. \tag{2.70}$$

2.1.5 Superposition of renewal processes

The superposition of independent renewal processes, or superimposed renewal process (SRP), is generally not a renewal process. An SRP can be determined by the times of events, $\{T_i\}$, and the source of events, $\{U_i\}$, where $U_i = k$ if the event that occurred at time T_i comes from the k-th source.

Initially employed in neurophysiology to model pulses or spikes from independent neurons received at the central nervous system [21, 26], or in queuing theory to model the inter-arrival times in a network of independent servers [113], an SRP describes, in reliability engineering, repairable series systems composed of independent units that undergo perfect and instantaneous repair: the failures in each unit form independent renewal processes, which result in an SRP when the outputs are pooled. Denote by λ_i the stochastic intensity of the i-th renewal process (which is a function of time, but we omit the subscript t for notational convenience). By virtue of the additive property of counting processes, the stochastic intensity of an SRP composed of n renewal processes is given by:

$$\lambda_s = \sum_{i=1}^{n}\lambda_i. \tag{2.71}$$

This simple relation explains why the superposition process is usually not renewal. In a renewal process, $\lambda_t = \lambda(B(t))$ where $B(t)$ stands for the age, as defined in Eq.(2.17). In an SRP, however, the process intensity at a given time t depends on the times elapsed since the last arrival from *each source*, i.e., $B_i(t)$, instead of the time elapsed since last arrival, i.e., $min(B_1(t), B_2(t)...B_n(t)$. Although the inter-arrival times in the superposition process are statistically dependent, it is straightforward to compute their distribution in the stationary state [77]:

$$R_s(t) = \lim_{k\to\infty} P(X_k \geq t) = \frac{\nu_1\nu_2\ldots\nu_n}{\nu_1+\nu_2+\ldots\nu_n}\sum_{i=1}^{n}R_i(t)\prod_{\substack{j=1\ldots n\\ j\neq i}}\Psi_j(t), \tag{2.72}$$

where n is the number of renewal processes, R_i is the survival function of the i-th renewal process, Ψ_i is defined as $\Psi_i(t) = \int_t^\infty R_i(u)\,du$ and ν_i, defined as $\nu_i = \frac{1}{\Psi_i(0)}$, is the equilibrium rate of the i-th renewal process, which corresponds to the inverse of the mean cycle duration. The joint distribution of m adjacent intervals has been discussed in [77], and

the special case where $m = 2$ is given by

$$R_s(s,t) = \lim_{k\to\infty} P(X_k \geq s, X_{k+1} \geq t) = \frac{\nu_1\nu_2\ldots\nu_p}{\nu_1+\nu_2+\ldots\nu_p}\sum_{i=1}^{p} R_i(s)R_i(t)\prod_{\substack{j=1\ldots p\\ j\neq i}}\Psi_j(s+t). \tag{2.73}$$

The behavior of an SRP in a non-stationary state has been considered in [11], and the renewal theory has been generalized for SRP by [71]. Industrial applications could be found in, among others, [87], [118], [88], [62] and [116] where the SRP has been used to model the failure/repair process for systems such as diesel engines, sockets of aero-engines, coating of steel structures and air-conditioning equipment.

2.1.6 Geometric Process

Definition 2.7 *Let $\{X_n, n \in \mathbb{N}^*\}$ be a sequence of independent random variables and $F_n(t)$ the Cdf of X_n. If*

$$F_n(t) = F(a^{n-1}t), \quad n = 1, 2, ... \tag{2.74}$$

where a is a positive constant, then the sequence $\{X_n\}$ is called a geometric process.

The geometric process (GP) is another generalization of a renewal process. First introduced by Smith & Leadbetter [102], the GP is used by Lam [72] to describe inter failure times with trends. The GP is referred to in [110] as a quasi-renewal process. A comprehensive theory with applications can be found in [73]. Another extension of GP has been considered in [12]. Obviously, if $a > 1$, $\{X_n\}$ is stochastically decreasing and converge to zero with probability 1; if $0 < a < 1$, $\{X_n\}$ is stochastically increasing and converge to infinity with probability 1; for $a = 1$, GP reduces to a renewal process. In reliability engineering, the case $a < 1$ is less common, as it suggests that the system is improving with each repair. Following Eq.(2.74),

$$E[X_n] = \frac{E[X_1]}{a^{n-1}}, \tag{2.75}$$

and

$$Var(X_n) = \frac{Var(X_1)}{a^{2(n-1)}}. \tag{2.76}$$

Furthermore, for $a > 1$, the sum of expected cycle durations is converging:

$$\sum_{n=1}^{\infty} E[X_n] = \frac{a}{a-1}E[X_1]. \tag{2.77}$$

The renewal-type equation for a geometric process is given by

$$M(t) = F(t) + \int_0^t M(a(t-x))dF(x). \tag{2.78}$$

It has been proved in [13] that for $a > 1$, $M(t)$ is infinite for all $t > 0$, which is not so evident: intuitively, $M(t)$ should be infinite when t is sufficiently large. More generalizations of the geometric process can be found in [14] and [41].

2.2 Heterogeneous population

Perfect homogeneity of objects is rare in nature and in the industry: it can be created in a laboratory, but not outside it. Due to the instability of production processes, environmental and other factors, most populations of manufactured items are heterogeneous. Nevertheless, most of the reliability modeling assume an homogeneous population. Ignoring heterogeneity can lead to serious errors in the reliability assessment of items and, as a consequence, to crucial economic losses.

In survival analysis, the term "frailty" instead of "heterogeneity" is more commonly used. It was suggested in [109] in a demographic context, based on the idea that heterogeneity in the human population might lead to differences in their mortality curve. The effects of frailty in survival analysis have been addressed in many key reference papers, including but not limited to [1], [2], [19]. As an extension of the proportional hazards model [20], frailty models could be classified as a univariate/multivariate frailty model, shared frailty model, or correlated frailty model [114]. In its simplest form, when no covariate is considered, frailty is an unobserved random proportionality factor, often denoted by Z, with realizations z, that modifies the failure rate of an individual, or of related individuals:

$$\lambda_i(t) = z_i \lambda(t). \tag{2.79}$$

Before introducing the frailty analysis for repairable systems and imperfect repair processes, let us have a closer look at how the heterogeneity influences the reliability of non-repairable systems.

2.2.1 Non-repairable system

The literature on heterogeneity in non-repairable systems and its influence is rich: the presence of unobserved heterogeneity will not only distort the shape of the failure/mortality rate curve [1] [44] [42], but will also influence the estimation of baseline distribution parameters. For example, Lindqvist [80] has observed that overlooking unobserved heterogeneity between monitored systems might lead to non-optimal, or completely erroneous, decisions. Heckman and Singer [54] have remarked that the parameter estimation is sensitive to the mixing distribution and proposed the use of nonparametric mixtures. Vallejos and Steel [106] have introduced a Bayesian inference on the mixture of Weibull survival distributions. Cha and Finkelstein [17] have presented several meaningful examples of models combining heterogeneity with reliability models. Here, for illustrative purposes, we present two typical cases: the mixture of two distributions and the continuous mixture.

Mixture of two distributions

Suppose that some good products are mixed with some defective ones. The time to failure of an item selected randomly from this population can be obviously described in terms of mixtures. Denote by $F_1(t)$ and $F_2(t)$, $f_1(t)$ and $f_2(t)$, $\lambda_1(t)$ and $\lambda_2(t)$ the Cdf, Pdf and failure rates of respectively the two subpopulations. Then, the mixture is defined by a mass π, representing the proportion of good products in the population. The mixture failure rate $\lambda_m(t)$ in this case, is given by

$$\lambda_m(t) = \frac{\pi f_1(t) + (1-\pi) f_2(t)}{\pi \overline{F}_1(t) + (1-\pi)\overline{F}_2(t)} \tag{2.80}$$

$$= \pi(t)\lambda_1(t) + (1-\pi(t))\lambda_2(t), \tag{2.81}$$

with

$$\pi(t) = \frac{\pi \overline{F}_1(t)}{\pi \overline{F}_1(t) + (1-\pi)\overline{F}_2(t)}. \tag{2.82}$$

Specifically, if $\lambda_1(t) \leq \lambda_2(t)$ for all $t > 0$, then

$$\lambda_1(t) \leq \lambda_m(t) \leq \lambda_2(t), \tag{2.83}$$

and $\pi(t)$ is increasing. The failure rate of the population converges therefore to $\lambda_1(t)$, the failure rate of the most "robust" item. This simply suggests that the most vulnerable items break down first. If both $\lambda_1(t)$ and $\lambda_2(t)$ are decreasing, i.e., DFR, then the mixture rate $\lambda_m(t)$ is also DFR [57]. However, the inverse is not true: the mixture of the population with IFR is not necessarily IFR. More information on specific shapes of the mixture failure rate of two distributions can be found in [53], [52], [8], [10] and [61].

Continuous mixture

Now, consider the situation where the frailty Z is a continuous random variable defined on $[0, \infty)$, with the Pdf $\pi(z)$. This is the case where all items in the population have different failure rates. Denote by $F(t|z)$, $f(t|z)$ and $\lambda(t|z)$ the conditional Cdf, Pdf and failure rate of an individual having frailty $Z = z$. Then, the conditional failure rate of an item drawn randomly from the population is given by

$$\lambda_m(t) = \frac{\int_0^\infty f(t|z)\pi(z)dz}{\int_0^\infty \overline{F}(t|z)\pi(z)dz}. \tag{2.84}$$

The most classic assumption on the mixture is the multiplicative model

$$\lambda(t|z) = z\lambda(t). \tag{2.85}$$

If the lifetime of the individuals are Weibull distributed with IFR, i.e., $\lambda(t) = \beta t^{\beta-1}$ with

$\beta > 1$, and that Z follows a gamma distribution with shape k and scale θ:

$$\pi(z|k,\theta) = \frac{1}{\Gamma(k)\theta^k} z^{k-1} e^{-\frac{z}{\theta}}, \tag{2.86}$$

then, Eq.(2.84) reduces to:

$$\lambda_m(t) = \frac{k\theta\beta t^{\beta-1}}{1+\theta t^{\beta}}. \tag{2.87}$$

Thus, $\lambda_m(t) = 0$ at $t = 0$, increases to its maximum at

$$t_{max} = \left(\frac{\beta-1}{\theta}\right)^{\frac{1}{\beta}}, \tag{2.88}$$

and converges to 0 as t tends to infinity. This gamma-distributed frailty model combined with the Weibull lifetime distribution has been studied in [107] to model the unobserved heterogeneity presented in safety valves from the North Sea. The shape of the mixture failure rate differs dramatically from the IFR $\lambda(t)$. Similar to the mixture of two distributions, this corresponds to the fact that the most vulnerable ones break down first: $\lambda_m(t)$ is initially increasing because the frail ones are failing, and is eventually decreasing because only the robust individual remains alive.

2.2.2 Repairable system

The literature on heterogeneity for repairable systems is less abundant than the one for non-repairable systems.. One of the early works is [37], where a compound Poisson process is used to model the behavior of repairable systems that have different failure intensities.

Minimal repair is then combined with the unobserved heterogeneity: Asfaw and Lindqvist [3] have investigated the heterogeneous population composed of independent NHPP using gamma-distributed frailty, while Slimacek and Lindqvist [100] have studied the parameter estimations in heterogeneous NHPP population when the distribution of frailty is unspecified. The results are then generalized in [101] to incorporate covariates. It should be mentioned that if the heterogeneity is erroneously ignored, the estimation of model parameters will be biased.

For illustrative purposes, consider m independent repairable systems that undergo minimal repair [3]. If the population is homogeneous, all individual systems share the same failure rate, $\lambda(t)$. Let the j-th systems be observed from time 0 to T_j. The i-th failure instant of the j-th system is denoted by $T_{j,i}$, with $j \in 1...m$ and $i \in 1...n_j$. Then, the likelihood of these data is given by

$$\mathcal{L} = \prod_{j=1}^{m} \left(\prod_{i=1}^{n_j} \lambda(T_{j,i}) \right) e^{-\Lambda(T_j)}. \tag{2.89}$$

Now, if the j-th individual is characterized by the failure rate $\lambda_j(t) = z_j\lambda(t)$, the likelihood

of this single system is given by:

$$\mathcal{L}(z_j) = \left(\prod_{i=1}^{n_j} z_j \lambda(T_{j,i}) \right) e^{-z_j \Lambda(T_j)}. \tag{2.90}$$

Denote by $\pi(z)$ the distribution for the frailty z. Since z_j is often unobservable, the contribution to the full likelihood from this system is obtained by conditioning with respect to z_j, i.e., compute the expected value:

$$\mathcal{L}_j = E[\mathcal{L}(z)] = \int_0^\infty \mathcal{L}(z)\pi(z)dz, \tag{2.91}$$

and the log-likelihood function of the population is obtained by summing up the individual log-likelihoods. The bias and standard deviations of the estimates when the model is not correctly specified have been thoroughly developed in [3], with the underlying power law failure intensity and gamma-distributed frailty. Maintenance optimization issues have been addressed in [98], where the authors investigated a fleet of repairable systems operating in different environments (covariates) by introducing a covariate-dependent trend renewal process and showed that the ignorance of the heterogeneity could lead to non-optimal maintenance plans as well as increased repair costs.

Chapter 3

Stable imperfect repair models: asymptotic properties

This chapter discusses relevant asymptotic properties for the steady-state virtual age processes. It is shown that the limiting distributions of age, the residual lifetime, and the spread that describe an ordinary renewal process can be generalized to the stable virtual age process, although the cycles of the latter are not independent. Asymptotic distributions of the virtual age at time t, as well as of the virtual ages at the start and the end of a cycle containing t (as t tends to infinity) are explicitly derived for two popular in practice imperfect maintenance models, namely, the Arithmetic Reduction of Age with infinite memory (ARA_∞) and the Brown-Proschan (BP) models. Some applications of the obtained results to maintenance optimization are discussed.

As shown in the previous chapter, ordinary renewal processes are stationary in the sense that the corresponding renewal density function is constant as time tends to infinity. The NHPP that describes minimal repairs is, obviously, non-stationary, and if, e.g., its rate is increasing, the failures are arriving more frequently with time. Kijima Type I and Geometric Process [79], like NHPP, are non-stationary and can be used to model the lifetime with trends. Although there are many publications on various applications of the virtual age models in reliability, not much has been done in the literature on the description of the relevant asymptotic properties of the corresponding virtual age processes.

It should be noted that the limiting properties of the ordinary renewal processes are especially important in various applications. For instance, obtaining the corresponding renewal functions can be computationally challenging, and simple asymptotic values provided by the renewal-type theorems are beneficial in practice. Another example is the alternating renewal process. The stationary availability in this case, which is usually of the main interest, is obtained simply via the mean up and downtimes of a system. The life cycles of many industrial systems are quite long, meaning that a large number of maintenance actions are performed. Moreover, in many instances, the operational data is recorded only when a system enters its stable regime. Therefore, the importance of asymptotic methods in the described context is hard to overestimate.

The study of asymptotic properties of the imperfect repair processes that more adequately

Notation	Interpretation	Expression
X_n	duration of the n-th cycle	$X_n = T_n - T_{n-1}$
A_n	virtual age after the n-th repair	
X_∞	duration of the asymptotic cycle	$X_\infty = \lim_{n\to\infty} X_n$
R_{X_∞}	limiting survival function of X_∞	$R_{X_\infty}(t) = P(X_\infty > t)$
f_{X_∞}	limiting Pdf of X_∞	$f_{X_\infty}(t) = -\frac{d}{dt} R_{X_\infty}(t)$
A_∞	asymptotic virtual age at the start of a cycle	$A_\infty = \lim_{n\to\infty} A_n$
R_{A_∞}	limiting survival function of A_∞	$R_{A_\infty}(t) = P(A_\infty > t)$
f_{A_∞}	limiting Pdf of A_∞	$f_{A_\infty}(t) = -\frac{d}{dt} R_{A_\infty}(t)$
μ	mean cycle duration in stationary state	$\mu = E(X_\infty) = \int_0^\infty R_{X_\infty}(x)\mathrm{d}x$
B_t	backward recurrence time at t	$B_t = t - T_{N_{t^-}}$
δ_t	remaining lifetime (forward recurrence time) at t	$\delta_t = T_{N_{t^-}+1} - t$
Y_t	spread at time t	$Y_t = B_t + \delta_t$
V_t^s	virtual age at the start of the cycle containing t	$V_t^s = A_{N_{t^-}}$
V_t	virtual age at time t	$V_t = A_{N_{t^-}} + t - T_{N_{t^-}}$
V_t^e	virtual age at the end of the cycle containing t	$V_t^e = V_t + \delta_t$

Table 3.1: Notations in stable Virtual Age models.

than ordinary renewal processes describe the maintenance of the real-world systems, seems to be *a natural and practically sound task* that is addressed in the current paper. For achieving this goal, we had to answer first the following questions: can asymptotic results for the age, the residual lifetime, and the spread for ordinary renewal processes be generalized (and under what conditions) to the case of the imperfect repair processes? What are the asymptotic distributions for these quantities? To answer these questions, specific theoretical results had to be obtained and illustrated by several practical examples.

The notations used in the rest of this section are gathered in Table 3.1, and the chapter is organized as follows: Section 3.1 presents the generalization of the limiting distributions of the age, the residual lifetime and the spread to the case of stable virtual age models; Section 3.2 discusses the distribution of virtual age; finally, some applications in maintenance optimization are addressed in Section 3.3 and concluding remarks are given in Section 3.4.

3.1 Asymptotic distributions of backward recurrence time, residual lifetime and spread in stable virtual age processes

Various generalizations of the 'standard renewal theory' were addressed in the literature in several publications. With relevance to our topic, the following papers (to name a few) can be of interest. For example, Chow and Robbins [18] considered the renewal theory in sequences with dependent and non-identically distributed intervals; Dagpunar [23] studied the renewal-type equations for a generalized Kijima type II process (see also [39] and [40]). Lam and Lehoczky [71] considered the generalizations of renewal theory to the superposition of renewal processes.

Our interest lies in the virtual age processes with cycles X_n converging in distribution to X_∞ as n tends to infinity. This property guarantees that the cycles are asymptotically identically distributed. The following theorem defines distributions of the backward recurrence time B_t, the residual lifetime δ_t and the spread Y_t for these *stable* virtual age processes.

Theorem 3.1 *In a stable virtual age processes with asymptotically identically distributed cycles, the limiting distributions of B_t and δ_t , similar to the standard renewal processes, are given by the following equilibrium distributions:*

$$\lim_{t\to\infty} P(B_t \le x) = \lim_{t\to\infty} P(\delta_t \le x) = \frac{1}{\mu}\int_0^x R_{X_\infty}(s)ds, \tag{3.1}$$

whereas the limiting distribution of the spread Y_t is

$$\lim_{t\to\infty} P(Y_t \le x) = \frac{1}{\mu}\int_0^x s\cdot f_{X_\infty}(s)ds. \tag{3.2}$$

proof 1 *Denote by $m'(t)$ the generalized renewal density function for the virtual age process. Thus, $m'(t)$ can be interpreted as the rate of the corresponding point process (similar to the 'standard' renewal density, which is the rate of the ordinary renewal process). Denote by $F(x,u)$ the Cdf of a cycle that had started at the calendar time u. Then, the Cdf of B_t, denoted by $F_{B_t}(x)$, can be written as the following integral*

$$F_{B_t}(x) = \begin{cases} \int_{t-x}^{t} \overline{F}(t-u,u)m'(u)du, & 0 \le x \le t \\ 1, & x > t, \end{cases} \tag{3.1.1}$$

where u is the time of the last repair before t and, accordingly, $m'(u)du$ is the probability that the cycle starts in $[u, u+du)$. The corresponding pdf is

$$f_{B_t}(x) = \begin{cases} \overline{F}(x,t-x)m'(t-x), & 0 \le x \le t \\ 0, & x > t. \end{cases} \tag{3.1.2}$$

Let x be fixed. Then for $t \to \infty$

$$\lim_{t\to\infty} f_{B_t}(x) = \frac{R_{X_\infty}(x)}{\mu}, \tag{3.1.3}$$

because

$$\overline{F}(x,t-x) \to_{t\to\infty} R_{X_\infty}(x), \quad \forall x \ge 0, \tag{3.1.4}$$

as the cycles converge in distribution to X_∞ and

$$m'(t-x) \to_{t\to\infty} \frac{1}{\mu}, \tag{3.1.5}$$

which results from the convergence of the cycles of the virtual age process and was shown in [23], Eq.(22) and in [18], Theorem 1.

The previous reasoning was for the backward recurrence time. A similar approach can be applied to the remaining lifetime δ_t and the spread Y_t. Thus, asymptotically, as $t \to \infty$, it is not necessary that the cycles are independent, as in the standard renewal theory, and it is sufficient that they are identically distributed. Moreover, in this case, the limiting distributions of B_t, δ_t and Y_t can be also proved similar to how it is elegantly performed in [95] using the corresponding alternating renewal process. The backward recurrence time is interpreted as the on-time, whereas the remaining lifetime is the off-time of the generalized alternating renewal process.

Due to Theorem 3.1 and, similar to the standard renewal theory, it holds asymptotically in our case that (the inspection paradox)

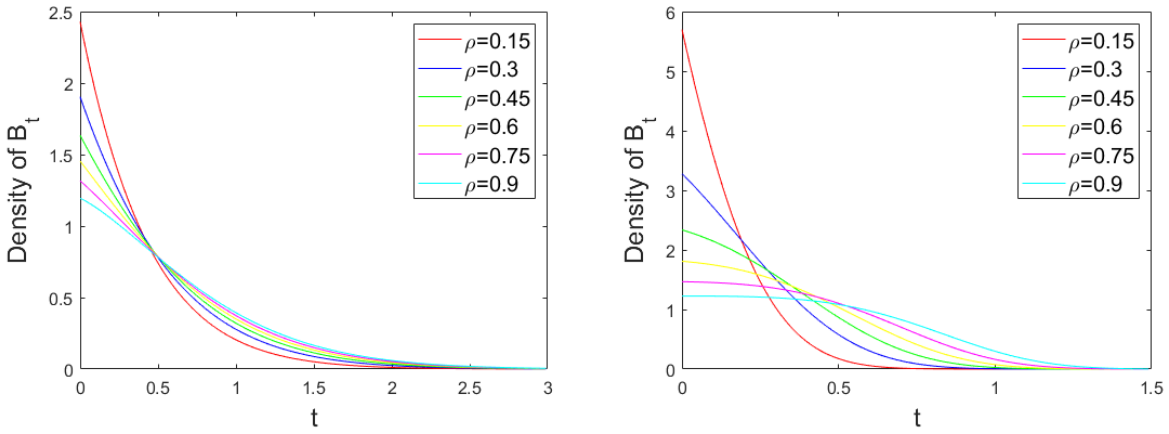
$$\forall s \in [0, \infty), \lim_{t\to\infty} R_{Y_t}(s) \geq R_{X_\infty}(s), \tag{3.3}$$

and

$$\lim_{t\to\infty} E[Y_t] = \frac{E[X_\infty^2]}{\mu}, \tag{3.4}$$

$$\lim_{t\to\infty} E[B_t] = \lim_{t\to\infty} E[\delta_t] = \frac{E[X_\infty^2]}{2\mu}. \tag{3.5}$$

The limiting distribution of X_∞ for the specific VA processes are given in the previous section, e.g., Eq.(2.68) for the Weibull ARA_∞ or Eq.(2.45) for the Brown Proschan process, whereas the existence of such distributions for a general ARA_∞ model is proved, e.g., in [39]. For illustration, the Pdf of the backward recurrence time B_t as t tends to infinity in an ARA_∞ process configured with $\alpha = 1$ are plotted in Figures 3.1a ($\beta = 1.5$) and 3.1b ($\beta = 4$). Clearly, the smaller ρ is, the more the Pdf of B_t is skewed towards the left.



(a) $\beta = 1.5$ (b) $\beta = 4$

Figure 3.1: Pdf of B_t when t tends to infinity in an ARA_∞.

3.2 Limiting distributions of V_t^s, V_t and V_t^e

As defined previously, $F(t)$ is the Cdf of the baseline distribution, and $F(t|a)$ is the Cdf of a cycle that starts with age a. The corresponding survival function is thus denoted

by $\overline{F}(t|a)$, whereas the mean residual life function is $\mu_a = \int_0^\infty \overline{F}(t|a)dt$. The Pdf of a cycle that starts with age a is denoted as $f(t|a) = F'(t|a)$. For stable VA processes (i.e., with asymptotically identically distributed cycles), it is important both from theoretical and practical points of view to obtain limiting distributions of the virtual age. This can be done for the two imperfect repair models, namely the ARA_∞ and Brown Proschan process.

3.2.1 ARA_∞ model

The limiting distributions of V_t^s, V_t and V_t^e for the ARA_∞ virtual age process are given by the following theorems.

Theorem 3.2 *Let $V_\infty^s = \lim_{t\to\infty} V_t^s$. The asymptotic Pdf of V_t^s is*

$$f_{V_\infty^s}(a) := \lim_{t\to\infty} f_{V_t^s}(a) = \frac{\mu_a}{\mu} \cdot f_{A_\infty}(a). \tag{3.6}$$

proof 2 *Conditioning the asymptotic distribution of B_t on V_t^s:*

$$\frac{1}{\mu}\int_0^y R_{X_\infty}(s)ds = \lim_{t\to\infty} P(B_t \le y) = \int_0^\infty \lim_{t\to\infty} P(B_t \le y|V_t^s = a) f_{V_t^s}(a)da. \tag{3.6.1}$$

Fixing a, we arrive at the standard renewal process. Therefore,

$$\lim_{t\to\infty} P(B_t \le y|V_t^s = a) = \frac{1}{\mu_a}\int_0^y \overline{F}(s|a)ds. \tag{3.6.2}$$

The left hand side of Eq.(3.6.1) can be alternatively expressed by conditioning on the virtual age at the start of a cycle:

$$\frac{1}{\mu}\int_0^y R_{X_\infty}(s)ds = \frac{1}{\mu}\int_0^y \int_0^\infty \overline{F}(s|a) f_{A_\infty} dads, \tag{3.6.3}$$

which results in

$$\frac{1}{\mu}\int_0^y \int_0^\infty \overline{F}(s|a) f_{A_\infty} dads = \int_0^\infty \int_0^y \frac{1}{\mu_a}\overline{F}(s|a) f_{V_\infty^s}(a)dsda. \tag{3.6.4}$$

Thus, obtaining the derivatives with respect to y,

$$\frac{1}{\mu}\int_0^\infty \overline{F}(y|a) f_{A_\infty}(a)da = \int_0^\infty \frac{1}{\mu_a}\overline{F}(y|a) f_{V_\infty^s}(a)da. \tag{3.6.5}$$

Therefore, an evident solution of the Pdf of V_∞^s is

$$f_{V_\infty^s}(a) = \frac{\mu_a}{\mu} \cdot f_{A_\infty}(a). \tag{3.6.6}$$

However, the equality in integral does not guarantee the equality in the integrand, so we need to prove the uniqueness of $f_{V_\infty^s}(a)$. Assume that a function $g(t)$ is the limiting Pdf of V_∞^s, with $\int_0^\infty g(t) = 1$ and

$$f_{V_\infty^s}(a) = g(a) \neq \frac{\mu_a}{\mu} \cdot f_{A_\infty}(a). \tag{3.6.7}$$

Reformulating Eq.(3.6.5) as:

$$\frac{1}{\mu} R_{X_\infty}(y) = \int_0^\infty \frac{1}{\mu_a} \overline{F}(y|a) g(a) da, \tag{3.6.8}$$

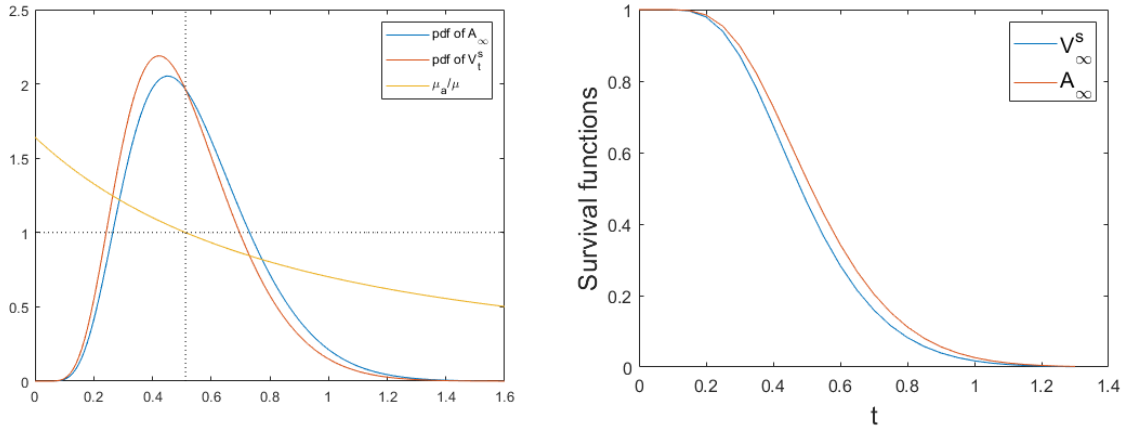
and considering the right hand side of above equation, we see that assuming $g(a) \neq \mu_a/\mu \cdot f_{A_\infty}(a)$ implies that the integrand could not be further simplified. Using the mean value theorem for integrals, there exists some $a^* \in [0, \infty)$ such that

$$\frac{1}{\mu} R_{X_\infty}(y) = \frac{1}{\mu_{a^*}} \overline{F}(y|a^*) \int_0^\infty g(a) da = \frac{1}{\mu_{a^*}} \overline{F}(y|a^*), \tag{3.6.9}$$

resulting in

$$\frac{\overline{F}(y|a^*)}{R_{X_\infty}(y)} = \frac{\mu_{a^*}}{\mu} = const. \tag{3.6.10}$$

This ratio cannot be a constant on all the points of $\mathbb{R}^+$ unless two survival functions are identical, which in our case, contradicts the assumption. Thus, the corresponding solution is unique, as given in Eq.(3.6).



(a) f_{A_∞}, $f_{V_\infty^s}$ and μ_a/μ (b) Survival functions

Figure 3.2: Example of Pdf and survival functions: $\alpha = 1$, $\beta = 2$, $\rho = 0.5$.

Interpretation of Eq.(3.6) depends on the baseline failure rate. For the IFR Weibull distribution, μ_a is a decreasing function of a and satisfies: $\mu_0 > \mu$ and $\lim_{a\to\infty} \mu_a = 0$. The Pdf of V_∞^s is formed by shifting f_{A_∞} to the left, and the two curves intersect at the point $a = \tilde{A}$ with $\mu_{\tilde{A}} = \mu$. An example is shown in Figure 3.2a.

A straightforward result of Theorem 3.2 is A_∞ being stochastically larger than $\lim_{t\to\infty} V_t^s$. It can be observed in Figure 3.2b that the survival function of A_∞ is constantly above that of V_∞^s. This means when a system that undergoes repairs of type ARA_∞ is observed

at time t with t tending to infinity, the repair just before t will make the system "younger" (with virtual age V_∞^s), compared to other repairs which reduce the virtual age to A_∞.

Consider now some aging notions related to stable states. Finkelstein [39] has defined the equilibrium age A^* that satisfies, if a cycle starts with age A^*, then the next cycle will also start with A^* but in expectation: $E[A_{i+1}|A_i = A^*] = A^*$. As defined previously, $\tilde{A}$ satisfies $\mu_{\tilde{A}} = \mu$, meaning that a cycle starting with age $\tilde{A}$ has the mean μ. Let $E[A_\infty]$ be the expected value of A_∞. Then, for the IFR baseline distributions,

$$\tilde{A} \leq A^* \leq E[A_\infty]. \tag{3.7}$$

proof 3 *Begin with the convexity of the function $g(a) = \mu_a$. $g(a)$ has derivative*

$$g'(a) = \int_0^\infty (-\lambda(t+a) + \lambda(a)) \frac{\overline{F}(t+a)}{\overline{F}(a)} dt. \tag{3.7.1}$$

This is an increasing function with $g'(0) = -1$ and $g'(\infty) = 0$. Thus, $g(a)$ is convex function.

Consider first A^* and $E[A_\infty]$. Let i tends to infinity. Since

$$E[A_{i+1}|A_i] = E[(1-\rho)(A_i + X_{i+1})] = (1-\rho)A_i + (1-\rho)E[X_{i+1}|A_i], \tag{3.7.2}$$

$E[A_{i+1}|A_i]$ is also a convex function of A_i. Therefore,

$$E[A_\infty] = E[E[A_{i+1}|A_i]] \geq E[A_{i+1}|A_i = E[A_\infty]]. \tag{3.7.3}$$

Let $h(A_i) = E[A_{i+1}|A_i] - A_i$. $h(A_i)$ is a decreasing function with $h(0) = (1-\rho)\mu_0$ and $h(\infty) = -\infty$. Obviously,

$$h(A^*) = E[A_{i+1}|A_i = A^*] - A^* = 0, \tag{3.7.4}$$

and because of Eq.(3.7.3),

$$h(E[A_\infty]) = E[A_{i+1}|A_i = E[A_\infty]] - E[A_\infty] \leq 0. \tag{3.7.5}$$

Thus, $E[A_\infty] \geq A^*$ given that h is decreasing.

Consider now A^* and $\tilde{A}$. Since $E[A_{i+1}A_i = A^*] = (1-\rho)A^* + (1-\rho)\mu_{A^*}$, we have

$$\mu_{A^*} = \frac{\rho}{1-\rho} A^*. \tag{3.7.6}$$

According to the definition of $\tilde{A}$, $\mu_{\tilde{A}} = \mu$. However, in the steady state of an ARA_∞ process, the following equality holds

$$\mu = \frac{\rho}{1-\rho} E[A_\infty], \tag{3.7.7}$$

leading to

$$\mu_{\tilde{A}} = \mu = \frac{\rho}{1-\rho} E[A_\infty] \geq \frac{\rho}{1-\rho} A^* = \mu_{A^*}. \tag{3.7.8}$$

$\tilde{A} \leq A^*$ *given that* μ_a *is a decreasing function of* a.

Theorem 3.3 *The limiting distributions of* V_t *and* V_t^e *are given, respectively, by*

$$\lim_{t\to\infty} P(V_t \leq y) = \frac{1}{\mu}\int_0^y \int_0^{y-a} \overline{F}(s|a) f_{A_\infty}(a) ds da, \tag{3.8}$$

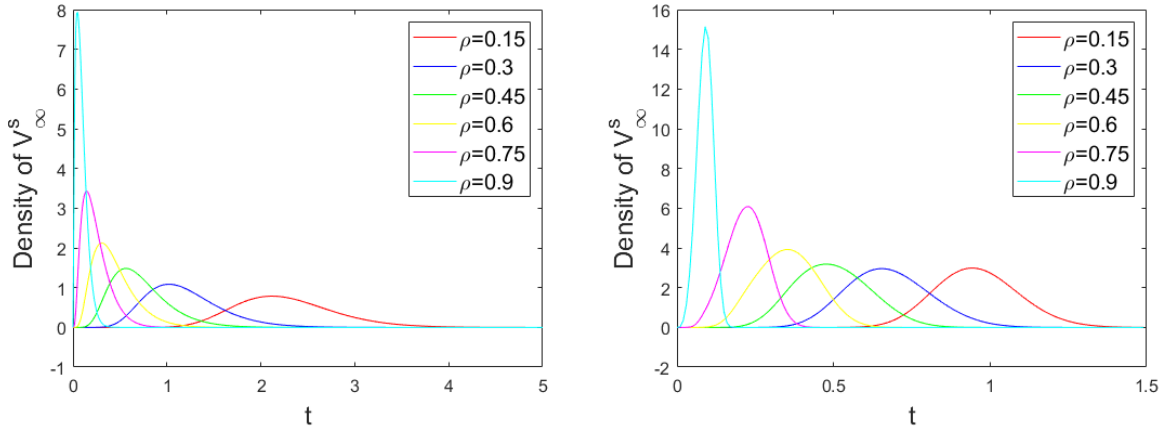
$$\lim_{t\to\infty} P(V_t^e \leq y) = \frac{1}{\mu}\int_0^y \int_0^{y-a} s \cdot f(s|a) f_{A_\infty}(a) ds da. \tag{3.9}$$

proof 4

$$\begin{aligned}
\lim_{t\to\infty} P(V_t \leq y) &= \int_0^y \lim_{t\to\infty} P(V_t \leq y | V_t^s = a) f_{V_t^s}(a) da \\
&= \int_0^y \lim_{t\to\infty} P(B_t \leq y-a | V_t^s = a) f_{V_t^s}(a) da \\
&= \int_0^y \frac{1}{\mu_a} \int_0^{y-a} \overline{F}(s|a) f_{V_t^s}(a) ds da \\
&= \frac{1}{\mu} \int_0^y \int_0^{y-a} \overline{F}(s|a) f_{A_\infty}(a) ds da.
\end{aligned}$$

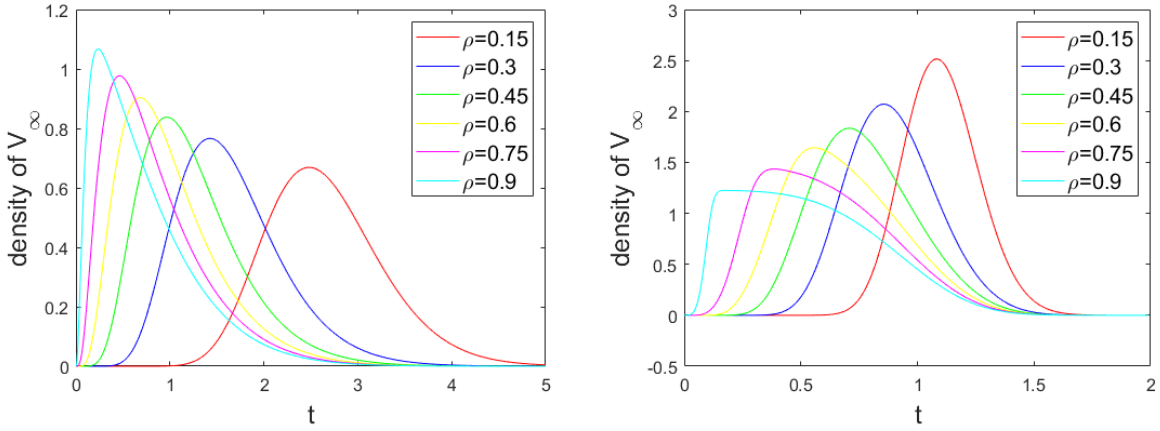
$$\begin{aligned}
\lim_{t\to\infty} P(V_t^e \leq y) &= \int_0^y \lim_{t\to\infty} P(V_t^e \leq y | V_t^s = a) f_{V_t^s}(a) da \\
&= \int_0^y \lim_{t\to\infty} P(Y_t \leq y-a | V_t^s = a) f_{V_t^s}(a) da \\
&= \int_0^y \frac{1}{\mu_a} \int_0^{y-a} s \cdot f(s|a) f_{V_t^s}(a) ds da \\
&= \frac{1}{\mu} \int_0^y \int_0^{y-a} s \cdot f(s|a) f_{A_\infty}(a) ds da.
\end{aligned}$$

Example 3.2.1 *Consider the* ARA_∞ *process with a power law baseline failure rate. Parameters are* $\alpha = 1$, β *and* ρ. *The Pdf of* V_t^s *and of* V_t *are given in Figures 3.3 and 3.4 for different values of* β *and of* ρ. *Figures 3.3a and 3.4a represent the case where* $\beta = 1.5$, *whereas Figures 3.3b and 3.4b stand for the case where* $\beta = 4$. *Unlike* B_t, V_t *and* V_t^s *are stochastically decreasing in* ρ: *the larger* ρ *is, the more the Pdf is skewed towards the left, which is comprehensible since repairs with a large* ρ *are more close to 'perfect', and that the ages after such repairs are stochastically 'smaller'.*



(a) $\beta = 1.5$ (b) $\beta = 4$

Figure 3.3: Pdf of V_t^s when t tends to infinity in an ARA_∞.



(a) $\beta = 1.5$ (b) $\beta = 4$

Figure 3.4: Pdf of V_t when t tends to infinity in an ARA_∞.

3.2.2 Brown-Proschan model

Distinct from the ARA_∞ model, the virtual age just after a repair in the BP process is not a continuous random variable, as it has a mass in the origin [76]. Therefore, our results formulated in Eqs.(3.6, 3.8, 3.9) cannot be directly applied, and the corresponding theorems should be proved differently. It should also be noted that it is practically important to obtain the mass for V_t^s in the origin, as it shows with what probability the last repair was perfect.

In accordance with our previous notation, let μ_0 be the mean duration of the first cycle, i.e., $\mu_0 = \int_0^\infty \overline{F}(t)dt$.

Theorem 3.4 *As t tends to infinity, V_t^s equals to 0 with probability*

$$\lim_{t\to\infty} P(V_t^s = 0) = \frac{p\mu_0}{\mu}, \tag{3.10}$$

and its density on $(0, \infty)$ is given by:

$$\lim_{t\to\infty} f_{V_t^s}(a) = \frac{\mu_a}{\mu} f_{A_\infty}(a). \tag{3.11}$$

proof 5 *The event $\{V_t^s = 0 \text{ as } t \to \infty\}$ is equivalent to $\{$the last repair before t as $t \to \infty$ is perfect$\}$. Let Tp be the waiting time between two perfect repairs in the corresponding BP process and F_p be its Cdf. By analogy with the alternating renewal process,*

$$\lim_{t\to\infty} P(V_t^s = 0) = \frac{E[X_1]}{E[T_p]} = \frac{\mu_0}{\int_0^\infty \overline{F}_p(s)ds}. \tag{3.10.1}$$

Following Lemma 2.1 in [15], $\overline{F}_p(t) = \overline{F}^p(t)$, leading to

$$\int_0^\infty \overline{F}_p(s)ds = \int_0^\infty e^{-p\Lambda(s)}ds = \frac{\mu}{p}, \tag{3.10.2}$$

which completes the proof of Eq.(3.10). To prove equation Eq.(3.11), condition the r.h.s. of Eq.(3.1) on V_t^s:

$$\lim_{t\to\infty} P(B_t \le y) = \lim_{t\to\infty} P(B_t \le y|V_t^s = 0)P(V_t^s = 0) + \lim_{t\to\infty} \int_{a=0^+}^{\infty} P(B_t \le y|V_t^s = a) f_{V_t^s}(a)da, \tag{3.11.1}$$

whereas conditioning of the l.h.s. on the virtual age at the start of a cycle results in

$$\frac{1}{\mu}\int_0^y R_{X_\infty}(s)ds = \frac{1}{\mu}\int_0^y \overline{F}(s|0)P(A_\infty = 0)ds + \frac{1}{\mu}\int_0^y \int_{a=0^+}^{\infty} \overline{F}(s|a) f_{A_\infty}(a)dads. \tag{3.11.2}$$

The first term of the r.h.s of Eq.(3.11.1) is equal to that of Eq.(3.11.2), because, using the corresponding alternating renewal process,

$$\lim_{t\to\infty} P(B_t \le y|V_t^s = 0) = \frac{E[min(y, X_1)]}{E[X_1]} = \frac{\int_0^y \overline{F}(s)ds}{\mu_0}, \tag{3.11.3}$$

thus,

$$\lim_{t\to\infty} P(B_t \le y|V_t^s = 0)P(V_t^s = 0) = \frac{p}{\mu}\int_0^y \overline{F}(s)ds, \tag{3.11.4}$$

which obviously equals to the first term of the r.h.s of Eq. (3.11.2). This also guarantees that the second terms of the r.h.s of Eqs. (3.11.1) and (3.11.2) are equal. Eq.(3.11) can therefore be proved in the same way as for the ARA_∞.

Corollary 3.2.1 *For the Weibull baseline distribution, $\mu = E[X_\infty] = p^{1-1/\beta}\mu_0$, which results in*

$$\lim_{t\to\infty} P(V_t^s = 0) = p^{\frac{1}{\beta}}. \tag{3.12}$$

Theorem 3.5 *The limiting distributions of V_t and V_t^e are respectively given by:*

$$\lim_{t\to\infty} P(V_t \le y) = \frac{p}{\mu}\int_0^y \overline{F}(s)ds + \frac{1}{\mu}\int_{0^+}^y \int_0^{y-a} \overline{F}(s|a) f_{A_\infty}(a)dsda, \tag{3.13}$$

$$\lim_{t\to\infty} P(V_t^e \le y) = \frac{p}{\mu}\int_0^y s\cdot f(s)ds + \frac{1}{\mu}\int_{0^+}^y \int_0^{y-a} s\cdot f(s|a) f_{A_\infty}(a)dsda. \tag{3.14}$$

proof 6 *Consider first V_t.*

$$\lim_{t\to\infty} P(V_t \le y) = \lim_{t\to\infty} P(V_t \le y|V_t^s = 0)P(V_t^s = 0) + \lim_{t\to\infty} \int_{0^+}^{y} P(V_t \le y|V_t^s = a) f_{V_t^s}(a)da. \tag{3.13.1}$$

The first term of the r.h.s. of Eq.(3.13.1) can be further developed as:

$$\begin{aligned}
&\lim_{t\to\infty} P(V_t \le y|V_t^s = 0)P(V_t^s = 0)\\
&= \lim_{t\to\infty} P(B_t \le y|V_t^s = 0)\lim_{t\to\infty} P(V_t^s = 0)\\
&= \frac{1}{\mu_0}\int_0^y \overline{F}(s)ds \cdot \frac{\mu_0}{\mu}p\\
&= \frac{p}{\mu}\int_0^y \overline{F}(s)ds,
\end{aligned}$$

whereas the second term of the r.h.s. of Eq.(3.13.1) can be derived in the same way as for the ARA_∞ model. Consider now V_t^e.

$$\lim_{t\to\infty} P(V_t^e \le y) = \lim_{t\to\infty} P(V_t^e \le y | V_t^s = 0)P(V_t^s = 0) + \int_{0^+}^{y} \lim_{t\to\infty} P(V_t^e \le y | V_t^s = a) f_{V_t^s}(a) da. \tag{3.14.1}$$

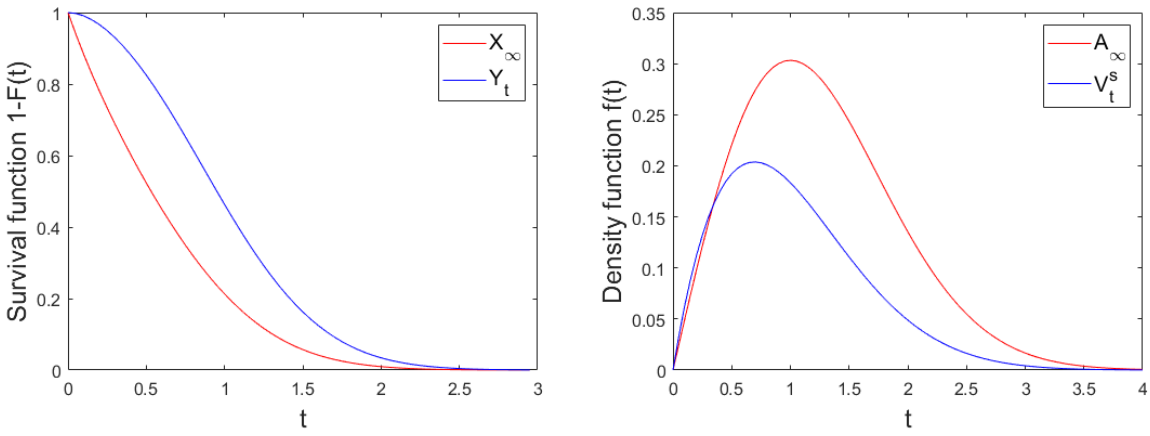
The first term of the r.h.s. of Eq.(3.14.1) can be further developed as:

$$\begin{aligned} &\lim_{t\to\infty} P(V_t^e \le y | V_t^s = 0)P(V_t^s = 0) \\ &= \lim_{t\to\infty} P(Y_t \le y | V_t^s = 0) \lim_{t\to\infty} P(V_t^s = 0) \\ &= \frac{1}{\mu_0}\int_0^y s \cdot f(s)ds \cdot \frac{p\mu_0}{\mu} \\ &= \frac{p}{\mu}\int_0^y s \cdot f(s)ds, \end{aligned}$$

whereas the second term of the r.h.s. of Eq.(3.14.1) can be derived in the same way as for the ARA_∞ model.

Example 3.2.2 *Consider the BP process with a power law baseline failure rate configured as $\alpha = 1$, $\beta = 2$, $p = 0.5$. The survival function of X_∞ (the asymptotic interval) and of $\lim_{t\to\infty} Y_t$ (the limiting spread) is plotted in Figure 3.5a. Since the blue curve is above the red one, Y_t is stochastically larger than X_∞, as stated in the inspection paradox. In Figure 3.5b, the Pdf of A_∞ (the asymptotic virtual age after a repair) and of $\lim_{t\to\infty} V_t^s$ (the limiting virtual age at the start of a cycle) on $(0, \infty)$ have been drawn with respectively the red curve and the blue curve. The mass on 0 is not shown in the figure. Obviously, the area under the red curve is significantly larger than that under the blue one. Therefore, the mass on 0 of $\lim_{t\to\infty} V_t^s$ is larger than that of A_∞. This is consistent with corollary 3.2.1: when $\beta > 1$, the probability that virtual age at the start of a cycle containing t as t tends to infinity is perfect, is larger than p.*

A physical interpretation is that, if we observe an intrinsically aging system that undergoes imperfect repair of type BP at an instant t long after the system begins to work, then the most recent repair has a higher probability of being perfect compared to a repair randomly drawn from the failure/repair process.



(a) Y_t vs X_∞ (b) A_∞ vs V_t^s

Figure 3.5: Y_t, X_∞, V_t^s and A_∞ in a BP process.

3.3 Application in maintenance: Optimal degree of imperfect repair

In this section, we focus mainly on the repair process of the ARA_∞- type with the increasing baseline failure rate. Finkelstein [43] has considered the optimal degree of imperfect repair that achieves the minimal, expected long-run cost rate for the repaired system accordingly. In the following, we shall make two extensions of this optimization problem. But first, let us recall the setting.

Assume that the cost of an imperfect maintenance action at any cycle depends only on the degree of repair ρ. Denote by $C(\rho)$ this cost. It is natural to assume that it is an increasing function of ρ and

$$C_m = C(0) \leq C(\rho) \leq C(1) = C_p, \tag{3.15}$$

where C_m and C_p are the costs of minimal and perfect repairs, respectively. Consider now the long-run average maintenance cost rate. The ARA_∞ process enters its steady-state and the mean cycle length for the corresponding Weibull IFR baseline distribution, $\mu(\rho) = E[X_\infty|\alpha, \beta, \rho]$, is an increasing function of ρ. Based on the renewal reward theory reasoning, the expected long-run cost per unit of time $c(\rho)$ is given by:

$$c(\rho) = \frac{C(\rho)}{\mu(\rho)}. \tag{3.16}$$

Assume a rather flexible functional form for $C(\rho)$

$$C(\rho) = C_m + (C_p - C_m)\rho^u, \quad u > 0. \tag{3.17}$$

Existence of an optimal maintenance degree ρ^*, which minimizes the long-run average cost rate $c(\rho)$ has been addressed in [43]. Basically, it requires that $c(\rho)$ be increasing as ρ tends to 1.

3.3.1 Recycling: reward based on backward recurrence time or on virtual age at retirement

The above reasoning considers the expected cost for an infinite horizon. In practice, however, systems are not operating forever, i.e., a 'retirement' threshold T_r is often predefined in a way that once the total working time exceeds T_r, system's operation is terminated (and it is usually replaced by a new one). The replaced system can sometimes be recycled, and the corresponding gain is generated following its condition, i.e., the better the condition, the larger the gain. A typical example is the garage of used cars, where the status of the used car and its accident history, are carefully examined to determine an appropriate price.

The optimization problem is formulated as follows: a system is under imperfect repair and is planned to 'retire' at time T_r. The expected lifetime in the steady-state regime,

$\mu(\rho)$ is a function of the repair degree ρ and satisfies $\mu(1) << T_r$. This is the condition guaranteeing that the system enters its steady-state before retiring. The cost of the repair actions is defined by Eq.(3.17). At time T_r, the system is recycled, and a reward, R_w, is assigned based on the state variable Y that can either be the backward recurrence time, B_{T_r} or the virtual age, V_{T_r}. Assume that the relation between the reward and Y is given by the following functional form:

$$R_w = re^{-\nu Y}, \quad r > 0, \nu > 0, \tag{3.18}$$

where r defines the maximal reward that could be obtained if the system is in the state "Good as New" at T_r. Under the condition $\mu(1) << T_r$, the expected long-run cost per unit of time can be defined, for instance, as

$$c_r(\rho) = \frac{C(\rho)}{\mu(\rho)} - \frac{r}{T_r} E[e^{-\nu Y}], \tag{3.19}$$

We are now interested in obtaining the optimal degree of repair ρ_r^* that minimizes the cost rate defined by Eq.(3.19). In practice, the corresponding decisions can be made either based on the virtual age of a system or on the elapsed time since the last repair (backward recurrence time). The latter, although giving less information on the state of a system, can be easier obtained, whereas V_t needs more information on the history of the repair process, which often can be unavailable. For ordinary renewal processes, these quantities are the same, whereas they are obviously different for the imperfect repair process.

3.3.1.1 Reward based on backward recurrence time: $Y = B_{T_r}$

Since $T_r >> \mu(1)$, the distribution of the backward recurrence time at retirement can be described by Eq.(3.1). The expected long-run cost rate is, therefore,

$$c_r(\rho|Y = B_{T_r}) = \frac{C(\rho)}{\mu(\rho)} - \frac{r}{T_r \mu(\rho)} \int_0^\infty e^{-\nu x} R_{X_\infty}(x|\rho) dx. \tag{3.20}$$

Consider the effect of a repair efficiency ρ on the expected value of the reward R_w. It can be seen from Definition 3.18 that this expected value for the IFR baseline distributions decreases when ρ is increasing (as the cycles of the corresponding steady-state virtual age process are stochastically increasing with ρ). Thus, given the parameters of the model, an optimal degree of repair can be found that minimizes Eq.(3.20). This is illustrated by the lowest curve in Figure (3.6).

3.3.1.2 Reward based on virtual age: $Y = V_{T_r}$

Assume now the reward is defined according to the virtual age V_{T_r} at the retirement time. As $T_r >> \mu(1)$, the virtual age tends to its asymptotic value and its distribution is given by Eq.(3.8). Given the Weibull baseline distribution, the expected long-run cost rate is

then defined as

$$c_r(\rho|Y = V_{T_r}) = \frac{C(\rho)}{\mu(\rho)} - \frac{r}{T_r\mu(\rho)} \int_0^\infty \int_0^x e^{-\nu x - \alpha x^\beta + \alpha a^\beta} f_{A_\infty}(a|\rho) da dx. \tag{3.21}$$

It is shown that the virtual age V_t as t tends to infinity is decreasing in ρ. Specifically, it tends to infinity when ρ tends to 0 (minimal repair) and then decreases to the asymptotic virtual age of the ordinary renewal process (which is just the corresponding backward recurrence time) when ρ tends to 1 (perfect repair). This behavior *dramatically differs* from that for B_{T_r} in the previous subsection (it was increasing in ρ), which is a meaningful fact. Eventually, it results in a larger optimal value of ρ than that defined by the cost rate function (Eq.(3.20)). Moreover, the optimal ρ for the case without recycling (Eq.(3.16)) lies between these two values, as is shown in Figure 3.6. The following numerical example illustrates our reasoning.

Example 3.3.1 *Let $\alpha = 1$, $\beta = 3$. Thus, the baseline survival function is $R(t) = e^{-t^3}$. Let $C_p = 1$, $C_m = 0.3$, $u = 4$. Then the long-run expected cost per unit of time without the recycling reward, is plotted by the solid line in Figure 3.6. The optimal repair degree $\rho^* \approx 0.57$ with a minimal expected cost $c(\rho^*) = 0.6966$.*

Consider now the reward policy defined as $T_r = 20$, $r = 20$, $\nu = 2$. The expected maximal reward defined by r/T_r equals one and is of the same order of magnitude as C_p. It is, therefore, necessary to take into account the reward when optimizing the maintenance degree. When the reward is based on the backward recurrence time at retirement, $\rho_r^* = 0.52$ with the corresponding expected cost rate $c(\rho_r^*) = 0.1363$ (dashed line) and when the reward is given according to the corresponding virtual age, then $\rho_r^* = 0.65$ with the expected cost rate $c(\rho_r^*) = 0.4195$ (dash-dotted line). These results are consistent with our previous analysis showing that B_t and V_t have an opposite impact on the value of ρ_r^*.

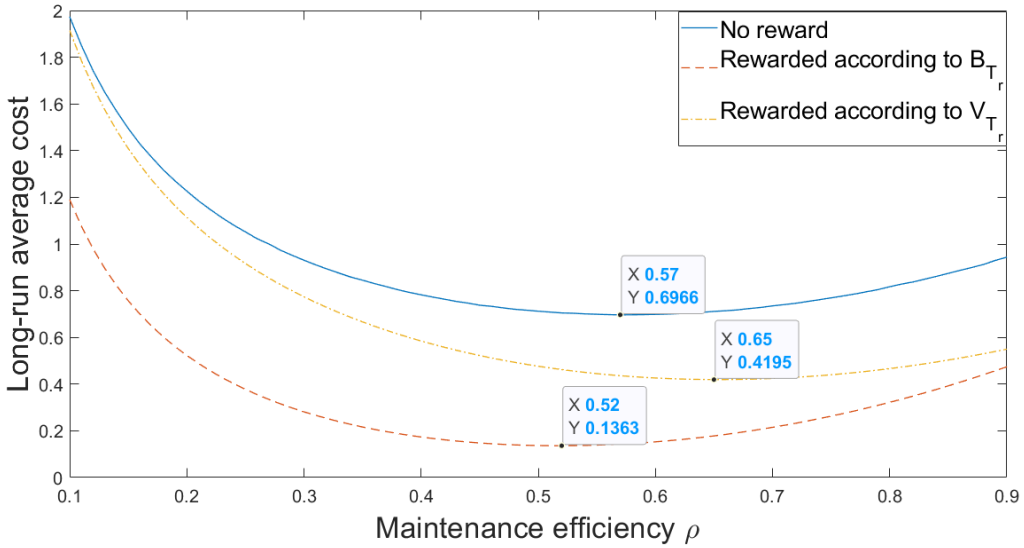


Figure 3.6: Optimal repair degrees.

3.3.2 Minimization of the long-run expected failure frequency of a series system under constrained budget

In this section, as another example, we deal with the constrained optimal imperfect maintenance problem [91] of the following type. We consider a series system of n independently operating and instantaneously maintained/repaired components with the imperfect repair of the ARA_∞ type. The first interval of the repair process of the *ith* component is Weibull-distributed with parameters (α_i, β_i), accordingly. The repair degrees of each component, $\vec{\rho} = \{\rho_1, \rho_2 ... \rho_n\}$ form the vector of decision variables. The repair cost, $C_i(\rho_i)$, depends only on the repair degree and is independent of the initial lifetime distribution. It is defined by Eq.(3.17) with different parameters for components. The expected long-run repair cost per unit of time of the system must not exceed the predefined cost threshold, C_{max}, i.e.,

$$\sum_{i=1}^{n} \frac{C_i(\rho_i)}{\mu(\alpha_i, \beta_i, \rho_i)} \leq C_{max}, \tag{3.22}$$

By the "long-run", as previously, we mean the steady-state case. Therefore, the denominator $\mu(\alpha_i, \beta_i, \rho_i)$ is the mean duration of the asymptotic cycle of the component i given α_i, β_i and ρ_i.

Under the constraint 3.22, we would like to minimize the steady-state failure frequency for the system. Thus, the corresponding objective function is defined as:

$$\lambda_s = \sum_{i=1}^{n} \frac{1}{\mu(\alpha_i, \beta_i, \rho_i)}. \tag{3.23}$$

In the following, for illustration, we will consider the simplest case of two components in series that differ only in the shape parameter β. Assume also that the cost function is the same for each component, i.e. $C_1(\rho) = C_2(\rho) = C(\rho)$.

Example 3.3.2 *Parameters of the Weibull distributions for the components are chosen as* $\alpha_1 = \alpha_2 = 1$, $\beta_1 = 1.5$, $\beta_2 = 3$, *whereas parameters of the cost function are:* $C_p = 1$, $C_m = 0.3$ *and* $u = 2$. $\vec{\rho} = (\rho_1, \rho_2)$ *is the decision vector. Obviously, without the constraint, the optimal repair degree is just* $\vec{\rho^*} = (1, 1)$. *However,when the maintenance cost threshold* C_{max} *is not large enough, we may not have enough resources to perform perfect repairs.*

The optimal repair degrees vector $\vec{\rho^*} = (\rho_1^*, \rho_2^*)$ *is defined by the points where the contour of the expected cost and that of the system's expected failure frequency are tangent to each other (Figure 3.7). When, for instance,* $C_{max} = [1.7, 1.8, 1.9, 2.0]$, *the corresponding optimal degrees are*

$$(\rho_1^*, \rho_2^*) = [(0.39, 0.67), (0.47, 0.89), (0.58, 0.98), (0.71, 1)],$$

accordingly, and the resulting minimal expected failure frequencies of the series system are $[3.36, 2.87, 2.62, 2.47]$. *Additionally, Figure 3.8 shows the corresponding pattern for the expected failure frequency of the system in the unconstrained case (no costs involved).*

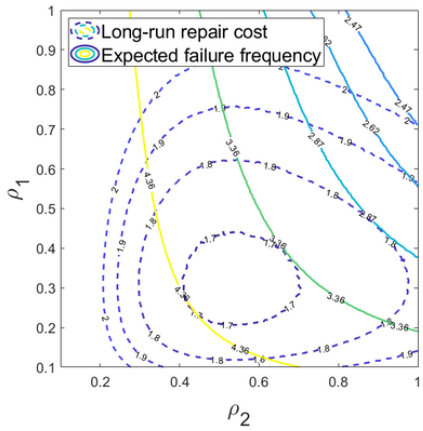


Figure 3.7: Optimal repair degrees determined by the contours that are tangent to each other.

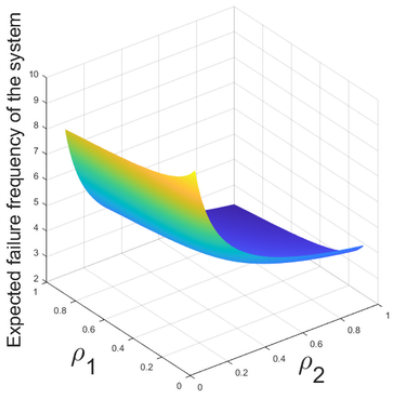


Figure 3.8: Expected failure frequency of the system as a function of the repair effectiveness.

3.4 Conclusion

This chapter studies asymptotic distributions for stable virtual age processes. We first show that the limiting distributions of the backward recurrence time, the remaining lifetime, and the spread that characterize an ordinary renewal process can be generalized to the case of the virtual age processes with asymptotically identically distributed cycles. Then we derive new analytical expressions for all limiting distributions of interest. We also discuss the importance of the age reduction mechanism for the obtained results. The provided examples highlight the practical value of our findings in reliability engineering.

This topic could be continued in the future in several directions. For instance, asymptotic distributions in stable virtual age models involving imperfect preventive maintenances can be considered. A typical example is the ARA_1CM-ARA_∞PM process described in [32]: corrective maintenances of the ARA_1 type are unable to keep the repaired system in a steady-state, whereas stationarity can be achieved by the periodic PMs of the ARA_∞ type. Therefore, it could be of interest to look at the asymptotic distribution of the virtual age just after the PM in this case. Limiting distributions in other imperfect maintenance models such as the Arithmetic reduction of intensity with infinite memory (ARI_∞) model [34] can also be worthy of further investigation.

Chapter 4

Imperfect repair models: heterogeneity

This chapter investigates the effect of heterogeneity on the failures of repairable systems that undergo imperfect repairs, which are extensively used in reliability engineering. When considering a group of similar systems, the assumption that the repair processes are independent and identically distributed becomes questionable owing to the unobserved heterogeneity in these systems. The basic models we consider include ARA_∞, ARA_1, Brown Proschan, and geometric process.

In Section 4.1, for the ARA_∞ process, we use the frailty model to study the proportional baseline hazard rate between the systems and, in particular, the gamma-distributed frailty is investigated. Thus, we derive the asymptotic properties of the mixed repair process and corresponding likelihood estimates and then evaluate the effects on the model parameter estimation when heterogeneity is erroneously ignored. Furthermore, when the model is established correctly by accounting for the gamma-distributed frailty, we find that the maximum likelihood estimator is inconsistent, and we propose an alternative approach. Two case studies are presented to illustrate the benefits of taking account of unobserved heterogeneity in scheduling preventive maintenance activities.

In Section 4.2, for the BP process, not only the proportional hazard rate is considered, but new parametric models have been proposed to describe the heterogeneous repair effectiveness and working environment as well. The impact of the heterogeneity on the mean lifetime of the population is revealed, and particularly, we highlight the importance of employing consistently adept maintenance crew.

In Section 4.3, for ARA_1 and geometric process, the influence of heterogeneous scale parameter have been addressed. We focus specifically on the parameter estimation issue. It is shown that when heterogeneity is erroneously overlooked, the aging parameter is constantly underestimated.

Finally, some applications of the relevant results in the analysis of the railway signaling system are presented in Chapter 6, showing the advantages of taking into account the heterogeneity when estimating the system's aging speed and reliability.

4.1 $\mathbf{ARA}_\infty$ model

We propose a multiplicative frailty model combined with ARA_∞-type imperfect maintenance. This is motivated by the observed decreasing failure rate of the railway signaling system, which contrasts with the common belief that electric components should have a constant or increasing failure rate (more details are given in Section 6). Some exchanges with experts in railway infrastructures lead us to consider individual heterogeneity, and to the intuition that a multiplicative frailty model could be a start because of its mathematical convenience.

4.1.1 Heterogeneous $\mathbf{ARA}_\infty$ population

In the context of frailty analyses, each individual in the population has its own intensity. This is modeled by $Z\lambda(t)$, whereby Z (mixing variable or frailty) is a non-negative random variable, and $\lambda(t)$ is the baseline intensity, which is common for all items. In general, it is assumed that the expectation of Z equals one, which makes $\lambda(t)$ an "average" intensity. There exist several potential distributions for Z, but the gamma distribution is the most commonly used one.

With the Weibull baseline distribution, $\lambda(t)$ is given by $\lambda(t) = \alpha\beta t^{\beta-1}$. Therefore, merging Z with α leads to a new pseudo scale parameter, $Z\alpha$. We address the influence of heterogeneity on the population without specifying the distribution of Z, before discussing the asymptotic properties and likelihood functions when Z is gamma-distributed.

4.1.1.1 Influence of unspecified frailty on the population mean lifetime

Let μ_Z be the mean inter-failure time of an ARA_∞ sequence with frailty Z:

$$\mu_Z = E[X_\infty | Z\alpha, \beta, \rho] = Z^{-1/\beta}\mu_0, \tag{4.1}$$

where $\mu_0 = E[X_\infty|\alpha, \beta, \rho]$. μ_Z is thus a random variable that represents the expected cycle durations in an individual ARA_∞ sequence given the frailty Z. $E[\mu_Z]$, on the other hand, can be regarded as the expected value of the duration of a cycle drawn from an ARA_∞ population when all the members have entered the stable regime. In the following of this section, $E[\mu_Z]$ is referred to as the *population mean lifetime*. The following proposition specifies its relationship with μ_0:

Proposition 1 *The population mean lifetime, $E[\mu_Z]$, if it exists, is larger than or equal to μ_0. Furthermore, if $\beta > 1$, $E[1/\mu_Z] \leq 1/\mu_0$.*

The proof is straightforward using Jensen's inequality, given that $E[Z] = 1$. Proposition 1 suggests that the frailty leads to a larger population mean lifetime, compared to the homogeneous case. Its application is shown later.

4.1.1.2 Survival function, mean and likelihood

We postulate the assumption that the pseudo scale parameter α follows a gamma distribution with a shape parameter k and scale parameter θ.

$$f_\alpha(s|k,\theta) = \frac{1}{\Gamma(k)\theta^k} s^{k-1} e^{-\frac{s}{\theta}}, \tag{4.2}$$

with mean $k\theta = \alpha$. This is equivalent to Z following a gamma distribution with mean 1. Thus, α is an "average" scale parameter to some degree. Consequently, an ARA_∞ population is fully determined by the quadruple (k, θ, β, ρ).

Let X^p_∞ be the steady-state population cycle duration. X^p_∞ does not represent any individual steady-state cycle duration but could be regarded as the duration of a cycle drawn from the ARA_∞ population when all the members have entered the stable regime. Similarly, let A^p_∞ be the steady-state population's virtual age, which is the VA after a repair of an item randomly drawn from the population that has entered the steady-state. The survival functions of A^p_∞ and X^p_∞, given parameters (k, θ, β, ρ), are obtained by conditioning R_{A_∞} and R_{X_∞} on the gamma distribution of α:

$$\begin{aligned} R_{A^p_\infty}(t|k,\theta,\beta,\rho) &= \int_0^\infty R_{A_\infty}(t|s,\beta,\rho) f_\alpha(s|k,\theta) ds \\ &= \sum_{s=1}^{\infty} \frac{1}{(q,q)_\infty (\frac{1}{q},\frac{1}{q})_{s-1}} (1 + \frac{\theta t^\beta}{q^s})^{-k}, \end{aligned} \tag{4.3}$$

$$\begin{aligned} R_{X^p_\infty}(t|k,\theta,\beta,\rho) &= \int_0^\infty R_{X_\infty}(t|s,\beta,\rho) f_\alpha(s|k,\theta) ds \\ &= \frac{\beta k}{\theta^k} \sum_{s=1}^{\infty} \frac{1}{q^s (q,q)_\infty (\frac{1}{q},\frac{1}{q})_{s-1}} \int_0^\infty x^{\beta-1} [(x+t)^\beta - (1-q^{-s})x^\beta + \frac{1}{\theta}]^{-(k+1)} dx. \end{aligned} \tag{4.4}$$

The population mean lifetime, $E[\mu_Z]$, and the necessary condition of $E[\mu_Z]$ being finite when α is gamma-distributed is shown in Proposition 2:

Proposition 2 *Let an ARA_∞ population be described by (k, θ, β, ρ) with $k\theta = \alpha$. $E[\mu_Z]$ is finite if and only if $k > \frac{1}{\beta}$. Let c_v be the coefficient of variation of α. As $c_v = k^{-1/2}$ in a gamma distribution, $E[\mu_Z] < \infty$ if and only if $c_v < \sqrt{\beta}$. When this condition is satisfied,*

$$E[\mu_Z] = \mu_0 k^{1/\beta} \frac{\Gamma(k - 1/\beta)}{\Gamma(k)}, \tag{4.5}$$

and

$$E[1/\mu_Z] = \frac{1}{\mu_0} k^{-1/\beta} \frac{\Gamma(k + 1/\beta)}{\Gamma(k)}, \tag{4.6}$$

The proof is straightforward and is omitted here.

To compute the likelihood function, begin with a single ARA_∞ sequence without considering the heterogeneity. Throughout Section 4.1, we use the notation

$$\mathcal{L}_\Delta^\Omega(\cdot|\cdot), \tag{4.7}$$

to represent the log-likelihood functions: $\Omega = s$ if the log-likelihood is derived for a *single* ARA_∞ process, and $\Omega = p$ for an ARA_∞ *population*; $\Delta = f$ if the pseudo scale parameter α is a constant (*fixed*), and $\Delta = r$ if α is a *random* variable because of the frailty.

In [89], the log-likelihood function based on a single ARA_∞ set of observations $\mathcal{X} = X_1, X_2...X_n$ is:

$$\mathcal{L}_f^s(\alpha, \beta, \rho|\mathcal{X}) = n \cdot log(\alpha\beta) + (\beta - 1)\sum_{i=1}^{n} log(a_{i-1} + X_i) - \alpha\sum_{i=1}^{n}(a_{i-1} + X_i)^\beta - a_{i-1}^\beta, \tag{4.8}$$

where a_{i-1} is the virtual age at the beginning of the i-th cycle. Consider now a homogeneous ARA_∞ population, i.e., all the systems are identical. Let $\mathcal{X}_M$ be the observation matrix where each row represents an individual ARA_∞ sequence of length n_j, whereby $j \in 1...M$. $X_{j,i}$ is the i-th interval in the j-th sequence, and $a_{j,i-1}$ is the virtual age at the beginning of the i-th cycle in the j-th sequence. The log-likelihood function is given by [32]:

$$\begin{aligned}\mathcal{L}_f^p(\alpha, \beta, \rho|\mathcal{X}_M) = log(\alpha\beta)\sum_{j=1}^{M} n_j + (\beta - 1)\sum_{j=1}^{M}\sum_{i=1}^{n_j} log(a_{j,i-1} + X_{j,i}) \\ - \alpha\sum_{j=1}^{M}\sum_{i=1}^{n_j}(a_{j,i-1} + X_{j,i})^\beta - a_{j,i-1}^\beta.\end{aligned} \tag{4.9}$$

Our interest is in the heterogeneous population. First, consider a single ARA_∞ process. When the pseudo scale parameter α is gamma distributed, the survival function of a cycle X starting at age v is given by:

$$R_{X|v}(t|\beta, k, \theta) = \int_0^\infty e^{-s((v+t)^\beta - v^\beta)} \cdot f_\alpha(s)ds = (\theta[(v + t)^\beta - v^\beta + 1/\theta])^{-k}, \tag{4.10}$$

with the Pdf and failure rate being expressed as,

$$f_{X|v}(t|\beta, k, \theta) = k\theta^{-k}\beta(v + t)^{\beta-1}[(v + t)^\beta - v^\beta + 1/\theta]^{-k-1}, \tag{4.11}$$

$$\lambda_{X|v}(t|\beta, k, \theta) = \frac{k\beta(v + t)^{\beta-1}}{(v + t)^\beta - v^\beta + 1/\theta}. \tag{4.12}$$

For a single observation sequence $\mathcal{X} = X_1, X_2...X_n$ with no censored data, the likelihood can be written as follows:

$$L(\beta, \rho, k, \theta|\mathcal{X}) = \prod_{i=1}^{n} k\theta^{-k}\beta(a_{i-1} + X_i)^{\beta-1}[(a_{i-1} + X_i)^\beta - a_{i-1}^\beta + \frac{1}{\theta}]^{-k-1}. \tag{4.13}$$

Therefore, its corresponding log-likelihood is:

$$\mathcal{L}_r^s(\beta,\rho,k,\theta|\mathcal{X}) = n\cdot log(k\theta^{-k}\beta)+(\beta-1)\sum_{i=1}^{n} log(a_{i-1}+X_i)-(k+1)\sum_{i=1}^{n} log((a_{i-1}+X_i)^{\beta}-a_{i-1}^{\beta}+\frac{1}{\theta}). \tag{4.14}$$

It should be noted that if only one ARA_∞ sequence is observed, the pair (k,θ) is non-identifiable because no information on the variation of α is available. Consider the case at which M independent ARA_∞ sequences are observed. Using the same notations, the log-likelihood of the ARA_∞ population becomes

$$\begin{aligned}\mathcal{L}_r^p(k,\theta,\beta,\rho|\mathcal{X}_M) = \sum_{j=1}^{M} n_j \cdot log(k\theta^{-k}\beta) + (\beta-1)\sum_{j=1}^{M}\sum_{i=1}^{n_j} log(a_{j,i-1}+X_{j,i}) \\ - (k+1)\sum_{j=1}^{M}\sum_{i=1}^{n_j} log((a_{j,i-1}+X_{j,i})^{\beta} - a_{j,i-1}^{\beta}+\frac{1}{\theta}).\end{aligned} \tag{4.15}$$

4.1.2 Inferences when heterogeneity is ignored

Given an observation matrix $\mathcal{X}_M$, it is important to properly specify the model if we want to estimate the parameters, which determines some reliability indicators for the system, e.g., the Weibull shape parameter β itself represents the aging speed of the asset. When the number of events of each system within a period of time does not differ from each other significantly, the heterogeneity is often overlooked. Under the faulty assumption that all the individual ARA_∞ sequences come from a certain triple (α, β, ρ), one may try to maximize the likelihood function defined by Eq.(4.9). In a statistical context, this is often referred to as the MLE of the misspecified model [112].

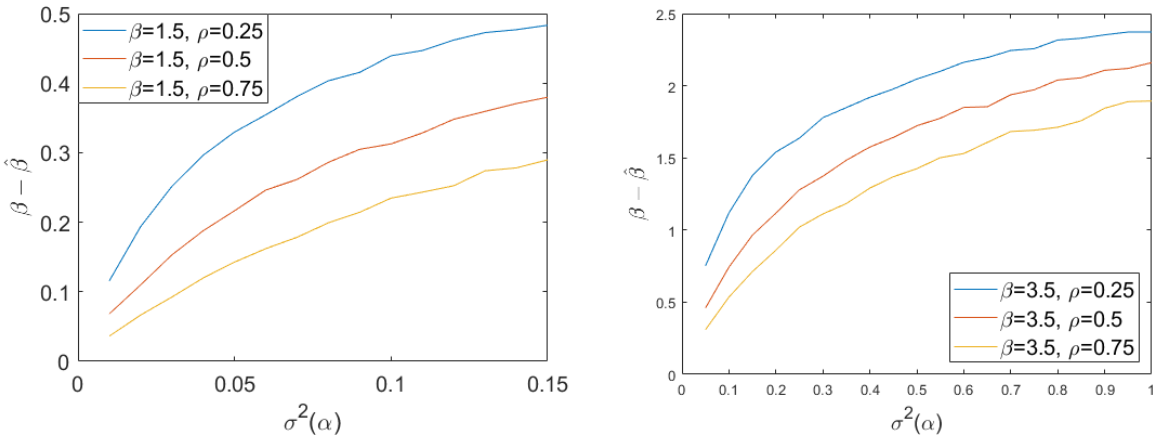
The following example shows how the variation of α influences the amplitude of the underestimation of β and overestimation of ρ. The ARA_∞ parameters are configured as follows: $\rho \in \{0.25, 0.5, 0.75\}$ for the cases associated with low/medium/high repair effectiveness, and $\beta \in \{1.5, 3.5\}$ associated the slow/fast wear out. In addition, α follows a gamma distribution $\Gamma(k,\theta)$, whereby the mean is fixed to unity: $k\theta = \overline{\alpha} = 1$. Let $\sigma^2(\alpha)$ be the variance of α. If the variation of α is considerably big, then the mean cycle duration $E[X_\infty]$ will also present considerable heterogeneity. Accordingly, maximization of the likelihood defined by Eq.(4.9) is obviously inappropriate. Thus, we will only investigate the situations at which the c_v value is *smaller than one*. For illustration, two vectors of $\sigma^2(\alpha)$ are considered: $\vec{\sigma^2}(\alpha) = 0.01 : 0.01 : 0.15$ for the case $\beta = 1.5$, and $\vec{\sigma^2}(\alpha) = 0.05 : 0.05 : 1$ for the case $\beta = 3.5$. The pseudocode of the Monte Carlo simulation is given below.

The length of a single ARA_∞ observation, N, was selected as 1000, and the total number of pooled ARA_∞, M, was set to 5000. These values were aimed at reducing the variance of the MLE estimators such that when the curves of the estimation bias are plotted, there should be some observable patterns or trends instead of points exhibiting large variations.

The estimation biases for the shape parameter β, defined by $D_\beta = \beta - \hat{\beta}$, are positive, plotted in Figures 4.1a and 4.1b. This indicates that β is underestimated when the

Algorithm 1 Measurement of the estimation bias when the inappropriate model is used

for $\beta = [1.5, 3.5]$ **do** *//aging rate*
  for $\rho = [0.25, 0.5, 0.75]$ **do** *//repair effectiveness*
    if $\beta = 1.5$ **then**
      $\vec{\sigma^2}(\alpha) = 0.01 : 0.01 : 0.15$
    else
      $\vec{\sigma^2}(\alpha) = 0.05 : 0.05 : 1$
    end if
    for $\sigma^2 \in \vec{\sigma^2}(\alpha)$ **do** *// variance of* α
      Determine k, θ: $\theta = \sigma^2, k = 1/\theta$
      Generate vector $\vec{\alpha}$ of length $M = 5000$ from $\Gamma(k, \theta)$
      for $i = 1 : M$ **do**
        Generate ARA_∞ sequence of length $N = 1000$
        with parameters $(\vec{\alpha}(i), \beta, \rho)$
      end for
      Construct observation matrix $\mathcal{X}_M$ of dim $M \times N$ where each line represents a single observation
      Formulate an inference on $\mathcal{X}_M$ using MLE (maximizing equation 4.9) and obtain $\hat{\beta}$ and $\hat{\rho}$
      Calculate bias: $D_\beta = \beta - \hat{\beta}$, $D_\rho = \rho - \hat{\rho}$
    end for
  end for
end for

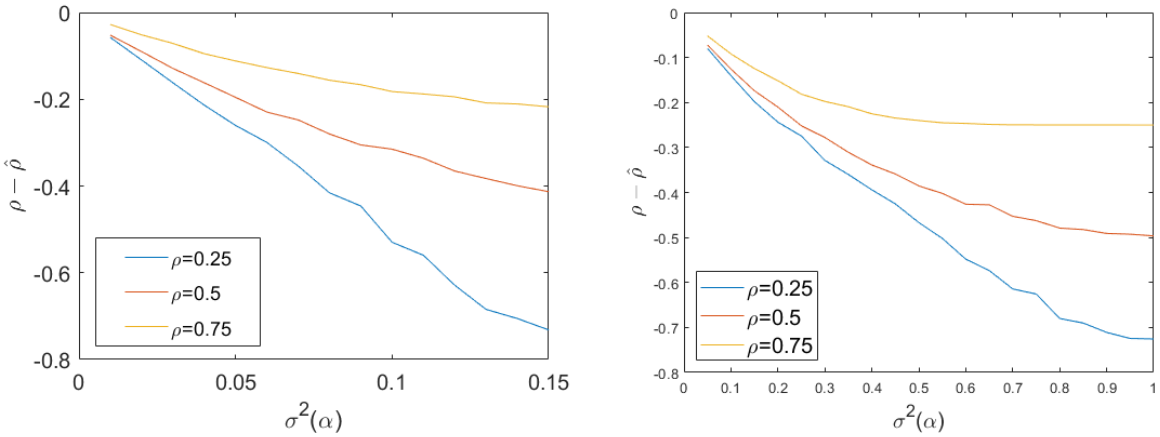


(a) Bias for $\hat{\beta}$, $\beta = 1.5$

(b) Bias for $\hat{\beta}$, $\beta = 3.5$

Figure 4.1: Bias for $\hat{\beta}$.

heterogeneity on α is overlooked. The amplitude of the underestimation of $\hat{\beta}$ (vertical axis) increases with the variation of α (horizontal axis). Blue, red and yellow curves represent the variation of estimation bias according to different repair efficiencies: the larger the ρ, the smaller the bias. Conversely, as shown in Figures 4.2a and 4.2b, $D_\rho = \rho - \hat{\rho}$ is negative, thus indicating that ρ is overestimated. Furthermore, the amplitude of under/overestimation increases with $\sigma^2(\alpha)$, and decreases when the repair effectiveness increases.



(a) Bias for $\hat{\rho}$, $\beta = 1.5$ (b) Bias for $\hat{\rho}$, $\beta = 3.5$

Figure 4.2: Bias for $\hat{\rho}$.

4.1.3 Inferences based on the correct model

Here we discuss the parameter estimation based on the correct model, i.e., we suppose that the population is heterogeneous, and the systems share the proportional hazard rates. Nevertheless, the problem is not as easy as it may look like: the traditional MLE method appears to be inconsistent, and some alternative inference procedure is proposed.

4.1.3.1 Maximum Likelihood Estimation

The ML estimators are obtained by maximizing Eq.(4.15):

$$(\hat{k}, \hat{\theta}, \hat{\beta}, \hat{\rho}) = \arg\max_{k,\theta,\beta,\rho} \mathcal{L}_r^p(k, \theta, \beta, \rho | \mathcal{X}_M). \tag{4.16}$$

The ML estimators are not consistent, that is, when the sample size tends to infinity, $\hat{k}, \hat{\theta}, \hat{\beta}$, and $\hat{\rho}$ do not converge to real parameter values. Nevertheless, the population mean lifetime (given by Eq.(4.5)) as well as the survival function (given by Eq.(4.4)) calculated with $\hat{k}, \hat{\theta}, \hat{\beta}$, and $\hat{\rho}$, do not differ considerably from those computed using real parameters. This is shown in the following example.

Example 4.1.1 *Consider the case where the ARA_∞ population is configured as follows: $k = 2$, $\theta = 0.5$, $\beta = 3.5$, and $\rho = 0.5$. The mean value of the scale parameter, $\overline{\alpha}$, equals one. To construct the observation matrix $\mathcal{X}_M$, generate $M = 10000$ independent ARA_∞ sequences, each with a length of $N = 1000$. The ML estimators obtained using Eq.(4.16) are: $\hat{k} = 6.3328$, $\hat{\theta} = 0.3859$, $\hat{\beta} = 2.072$, and $\hat{\rho} = 0.8471$. Thus, although the likelihood is correctly established, β is still underestimated, whereas ρ is overestimated.*

Underestimation of β and overestimation of ρ causes overestimation of the reliability of the item. In our case, the reliability is not overestimated because the average value of the scale parameter following the gamma distribution with k and θ was overestimated: $\hat{\overline{\alpha}} = \hat{k}\hat{\theta} = 2.4437 > 1 = \overline{\alpha}$. Thus, the mean cycle durations of the population are very close,

that is, $E[X^p_\infty|k,\theta,\beta,\rho] = 0.5263$*, and* $E[X^p_\infty|\hat{k},\hat{\theta},\hat{\beta},\hat{\rho}] = 0.5268$*. When the sample size tends to infinity, the population mean lifetime calculated with ML estimators, is observed to converge to the that computed with the true parameters. This is also observed for other configurations that we tested. Besides, the distance between the survival functions,* $R_{X^p_\infty}(t|k,\theta,\beta,\rho)$ *and* $R_{X^p_\infty}(t|\hat{k},\hat{\theta},\hat{\beta},\hat{\rho})$*, is also almost negligible.*

Although, in Example 4.1.1, the population mean lifetime is nearly unbiased when the model is erroneously specified, the consequences of a biased estimator are highlighted later in maintenance optimization. We present in the following an alternative estimator, which is asymptotically consistent, i.e., converge to the true parameter values when the sample size tends to infinity.

4.1.3.2 Alternative approach

The principle on which our alternative approach is based is that the estimation of β and ρ could be achieved independently from α. This is illustrated by the fact that when only one ARA_∞ sequence is observed, β and ρ could be accurately estimated if the failure data is sufficiently long. Thus, in our proposed alternative approach, we first estimate β and ρ and then deduce k and θ.

Estimation of β and ρ

Consider the observation matrix $\mathcal{X}_M$ in which the j-th row represents a single ARA_∞ observation with a length n_j: $\mathcal{X}_j = X_{j,1}, X_{j,2}...X_{j,n_j}$. Given β and ρ, the scale parameter α_j can be estimated by setting the derivative of the likelihood function (Eq.(4.8)) with respect to α_j to zero:

$$\hat{\alpha_j}(\beta,\rho) = \frac{n_j}{\sum_{i=1}^{n_j}(a_{j,i-1}+X_{j,i})^\beta - a^\beta_{j,i-1}}. \tag{4.17}$$

This means that we can estimate individually the pseudo parameters α_j. The likelihood of a single ARA_∞ sequence given β and ρ is thus $\mathcal{L}^s_f(\hat{\alpha_j}(\beta,\rho),\beta,\rho|\mathcal{X}_j)$. When M independent ARA_∞ sequences are superimposed in total, the likelihood given β and ρ is simply the sum of all individual log-likelihoods:

$$\begin{aligned}\mathcal{L}^*(\beta,\rho|\mathcal{X}_M) = \sum_{j=1}^{M} n_j \cdot log(\hat{\alpha_j}(\beta,\rho)\beta) + (\beta-1)\sum_{j=1}^{M}\sum_{i=1}^{n_j} log(a_{j,i-1}+X_{j,i}) \\ - \sum_{j=1}^{M}\hat{\alpha_j}(\beta,\rho)\sum_{i=1}^{n_j}(a_{j,i-1}+X_{j,i})^\beta - a^\beta_{j,i-1}.\end{aligned} \tag{4.18}$$

The alternative estimators of β and ρ, denoted as β^* and ρ^*, are given as follows:

$$(\beta^*,\rho^*) = \arg\max_{\beta,\rho} \mathcal{L}^*(\beta,\rho|\mathcal{X}_M). \tag{4.19}$$

Estimation of k and θ

Two possible measures are required to evaluate k and θ once β and ρ have been estimated. The first among them is to derive the individual scale parameters $\alpha_j, j \in 1...M$ using β^*, ρ^*:

$$\alpha_j^* = \frac{n_j}{\sum_{i=1}^{n_j}(a_{j,i-1} + X_{j,i})^{\beta^*} - (a_{j,i-1})^{\beta^*}}. \tag{4.20}$$

$\alpha_j^*, j \in 1...M$ are independently and identically distributed (i.i.d.) random variables following a gamma distribution with parameter (k_a^*, θ_a^*), which could be estimated from the standard MLE for gamma distribution. An alternative approach is to estimate k and θ by maximizing the population likelihood function in Eq.(4.15). The derivatives of Eq.(4.15) with respect to k and θ are:

$$\frac{\partial \mathcal{L}_r^p(k,\theta,\beta,\rho|\mathcal{X}_M)}{\partial k} = \frac{1}{k}\sum_{j=1}^{M} n_j - log(\theta)\sum_{j=1}^{M} n_j - \sum_{j=1}^{M}\sum_{i=1}^{n_j} log((a_{j,i-1} + X_{j,i})^{\beta} - a_{j,i-1}^{\beta} + \frac{1}{\theta}), \tag{4.21}$$

$$\frac{\partial \mathcal{L}_r^p(k,\theta,\beta,\rho|\mathcal{X}_M)}{\partial \theta} = -\frac{k}{\theta}\sum_{j=1}^{M} n_j + (k+1)\sum_{j=1}^{M}\sum_{i=1}^{n_j} \frac{1}{\theta^2(a_{j,i-1} + X_{j,i})^{\beta} - (a_{j,i-1})^{\beta} + \frac{1}{\theta}}. \tag{4.22}$$

θ_b^* and k_b^* can in turn be numerically obtained by setting the derivatives to zero,

$$\begin{cases} \dfrac{\partial \mathcal{L}_r^p(k,\theta,\beta^*,\rho^*|\mathcal{X}_M)}{\partial k}\Big|_{k=k_b^*,\theta=\theta_b^*} = 0 \\ \dfrac{\partial \mathcal{L}_r^p(k,\theta,\beta^*,\rho^*|\mathcal{X}_M)}{\partial \theta}\Big|_{k=k_b^*,\theta=\theta_b^*} = 0 \end{cases} \tag{4.23}$$

(k_a^*, θ_a^*) and (k_b^*, θ_b^*) are compared in the next section.

4.1.4 Bias and variance of the alternative estimator

We address the consistency of the proposed alternative estimator when the sample size is finite or infinite. For illustrative purpose, we study the ARA_∞ population configured as $\beta = 2$, $\rho = 0.75$, $k = 4$, $\theta = 0.25$. The individual ARA_∞ sequences share the common length: $n_1 = n_2 = ... = n_M = N$.

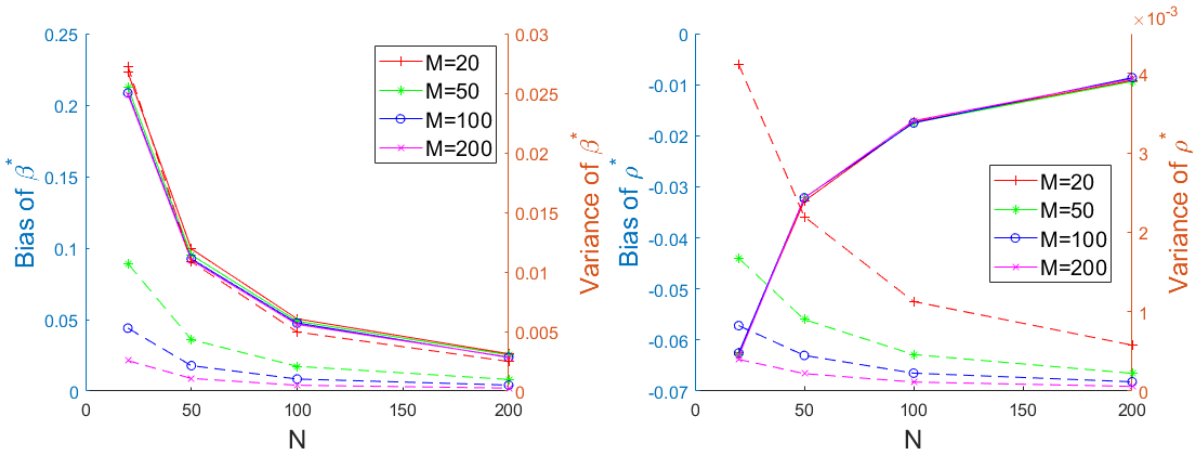
The asymptotic consistency of the estimators is demonstrated by estimating the parameters from a very large sample (generated by Monte Carlo simulation): setting $N = 20000$ and $M = 20000$, the parameters are estimated as: $\beta^* = 2.0004$, $\rho^* = 0.7499$, $k^* = 3.9914$ and $\theta^* = 0.2508$. In practice, however, the "infinite sample" can never be achieved, which makes it necessary to investigate how the bias and variance of the estimates change as a function of N and M.

4.1.4.1 β^* and ρ^*

The combinations of $M \in \{20, 50, 100, 200\}$ and $N \in \{20, 50, 100, 200\}$ are studied. A total of 5000 independent samples, each of size $M \times N$, are generated. 5000 estimates of the parameters, $\beta_1^*, \beta_2^*, ...\beta_{5000}^*$ and $\rho_1^*, \rho_2^*, ...\rho_{5000}^*$ are obtained for a given M and N. The bias for β^* is defined as $1/5000 \sum_i \beta_i^* - \beta$ (the same goes for ρ).

On the one hand, the bias for β^* is positive, decreasing in N (solid lines in Figure 4.3a), whereas the bias for ρ^* (solid lines in Figure 4.3b) is negative and its absolute value is also decreasing in N. M has, however, no influence on the bias, which means that the accuracy of the estimates is not improved by increasing the number of independent ARA_∞ sequences. We conclude that the alternative estimators over-estimates β and under-estimates ρ. On the other hand, their variances are decreasing in both N and M (dashed lines in Figures 4.3a and 4.3b).

Compare now β^* and ρ^* to the maximum likelihood estimators, $\hat{\beta}$ and $\hat{\rho}$. For the configuration studied above, the bias for $\hat{\beta}$ ranges from -0.47 to -0.55, whereas that of $\hat{\rho}$ ranges from 0.18 to 0.21. MLE estimators are, therefore, much more biased (with finite or infinite sample) than the alternative estimators.



(a) β^* (b) ρ^*

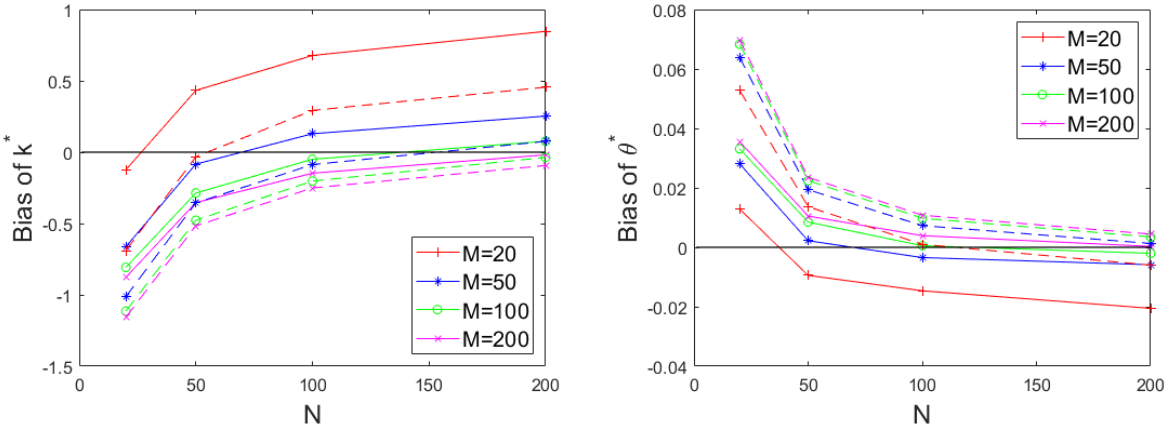
Figure 4.3: Bias and variance of β^* (left) and of ρ^* (right). The y-axis on the left side represents the bias (drawn correspondingly in the graphs with solid lines) while that on the right side shows the variance (dashed lines).

4.1.4.2 k^* and θ^*

The accuracy of the estimators k^* and θ^* relies on 1) an accurate estimation of β and ρ and 2) a sufficient number of independent ARA_∞ sequences. Figures 4.4a and 4.4b show that the estimates of k and θ, no matter obtained with gamma fit (k_a^* and θ_a^*) or with likelihood maximization (k_b^* and θ_b^*), converge to k and θ only if both M and N are very large (the pink line with cross marker, representing the bias when $M = 200$, is the only one that converge to 0 (the black line) as N grows). This being, the practitioner who wishes to estimate k and θ needs to choose the appropriate inference method *according to the data size*: for example, when a total of 20 ARA_∞ sequences are observed, using

the gamma fit approach is more suitable than utilizing the likelihood maximization when a large number of observations is recorded for each individual sequence (when $N = 200$, dashed red lines are closer to 0 compared to solid red lines); if, however, the numbers of observations in each ARA_∞ sequence are also limited, e.g., 20, then the ML estimators, k_b^* and θ_b^*, becomes preferable than k_a^* and θ_a^*.

Besides, the variance of $k_{a/b}^*$ and $\theta_{a/b}^*$ is not influenced by N but decreases when M increases.



(a) k_a^* and k_b^* (b) θ_a^* and θ_b^*

Figure 4.4: Bias for k^* (left) and for θ^* (right). k_a^* and θ_a^* are plotted using dashed lines: they are derived by fitting a gamma distribution to α. k_b^* and θ_b^* are plotted using solid lines: they are obtained by maximizing the corresponding likelihood.

4.1.5 Maintenance optimization for heterogeneous ARA_∞ population

We consider here two typical maintenance optimization problems. First, the issue of optimal repair degree when the cost of corrective maintenance is an increasing function of ρ is examined for heterogeneous ARA_∞ population. Second, the block replacement PM policy is investigated. We show that the heterogeneity does not influence the optimal repair degree, but affects the long-run cost per unit of time per system and the optimal PM interval. Particularly, Example 4.1.3 illustrates the benefits of taking the heterogeneity into account when scheduling the periodic PM with unknown parameters.

4.1.5.1 Optimal repair degree

In this section, we focus mainly on the repair process of heterogeneous ARA_∞ population with the increasing baseline failure rate. Finkelstein [43] has considered the optimal degree of imperfect repair that achieves the minimal, expected long-run cost rate for one single system. In the following, we investigate how the frailty, not necessarily gamma-distributed, influence the expected long-run cost rate as well as the optimal repair degree. But first, let us recall the setting.

Assume that the degree of repair, ρ, is a decision variable and that the cost of an imperfect maintenance action at any cycle depends only on ρ. This situation occurs when one can decide to what extent the system is maintained. For example, let a series system be composed of many independent components that are aging. When one component fails, one may choose to 1) replace the failed item (which almost corresponds to a minimal repair since only a small proportion of the system is renewed) or 2) replace all (a perfect repair) or 3) replace a certain number of components, including the failed one. The third case corresponds to an imperfect repair at a system level, and the proportion of the replaced items determines ρ. Denote by $C(\rho|\alpha,\beta)$ this cost. It is natural to assume that it is an increasing function of ρ and

$$C_m = C(0|\alpha,\beta) \leq C(\rho|\alpha,\beta) \leq C(1|\alpha,\beta) = C_p, \tag{4.24}$$

where C_m and C_p are the costs of minimal and perfect repairs, respectively. Consider now the long-run average maintenance cost rate. The ARA_∞ process enters its steady-state and the mean cycle length for the corresponding Weibull IFR baseline distribution, $\mu(\rho|\alpha,\beta) = E[X_\infty|\alpha,\beta,\rho]$, is an increasing function of ρ. Based on the renewal reward theory reasoning, for a single system, the expected long-run cost per unit of time $c(\rho)$ is given by:

$$c(\rho|\alpha,\beta) = \frac{C(\rho|\alpha,\beta)}{\mu(\rho|\alpha,\beta)}. \tag{4.25}$$

This is also the long-run average repair cost per unit of time per system when maintenance activities are carried out on a homogeneous population. Existence of an optimal maintenance degree ρ^*, which minimizes $c(\rho|\alpha,\beta)$, has been addressed in [43]. Basically, it requires that $c(\rho|\alpha,\beta)$ be increasing as ρ tends to 1.

Consider now a heterogeneous ARA_∞ population. The baseline failure intensity of the systems is characterized by frailty Z. The corresponding long-run average cost rate is therefore given by:

$$c(\rho|Z\alpha,\beta) = \frac{C(\rho|\alpha,\beta)}{\mu_Z(\rho|\alpha,\beta)}, \tag{4.26}$$

where $\mu_Z(\rho|\alpha,\beta) = E[X_\infty|Z\alpha,\beta,\rho]$. The long-run average cost rate per system can be obtained by taking the expectation:

$$E[c(\rho|Z\alpha,\beta)] = C(\rho|\alpha,\beta)E[\frac{1}{\mu_Z(\rho|\alpha,\beta)}] \leq c(\rho|\alpha,\beta). \tag{4.27}$$

The inequality follows Proposition 1. It seems that the frailty reduces the long-run average cost per unit of time per system. This is because the systems in a heterogeneous population survive, on average, "longer" than those in a homogeneous one. In addition, it is obvious that the optimal repair degree ρ^* minimizing $c(\rho|\alpha,\beta)$ will also minimize $E[c(\rho|Z\alpha,\beta)]$. We, therefore, conclude that the frailty reduces the long-run average cost rate per system without influencing the optimal repair degree.

4.1.5.2 Block replacement policy in ARA_∞ population

The block replacement policy [67] is widely used in industry because it is easier to administer than other policies that are potentially more cost-effective, like age-based PM or failure-limit PM [89]. It consists of replacing the item periodically at a fixed interval of $\tau, 2\tau, 3\tau...$, and the periodic replacement is supposed to be perfect in the current study. Here, we make the assumption that if an item fails within an interval, it is imperfectly repaired with a repair degree ρ. There is no maintenance delay, and the repair time is assumed to be negligible.

For a single ARA_∞ process, the long-run average cost per unit of time is given by:

$$c(\tau|\alpha,\beta,\rho) = \frac{C_p + C_c \cdot M(\tau, 0|\alpha,\beta,\rho)}{\tau}, \tag{4.28}$$

where C_p and C_c are respectively the cost of PM and of CM. $M(t, v|\alpha,\beta,\rho)$ is the renewal-type function in an ARA_∞ process with parameters (α,β,ρ), representing the expected value of number of events within an interval t when the system starts working at age v. It is the solution to the following renewal-type equation:

$$M(t, v|\alpha,\beta,\rho) = \int_0^t [1 + M(t-y, (1-\rho)(v+y)|\alpha,\beta,\rho)] \frac{f(v+y|\alpha,\beta,\rho)}{\overline{F}(v|\alpha,\beta,\rho)} dy, \tag{4.29}$$

where f and $\overline{F}$ are the Pdf and survival function of the first interval (see [23] for more details). Eq.(4.28) holds for a homogeneous population, wherein each item shares the same PM period. Now, let us consider a heterogeneous ARA_∞ population governed by the frailty Z. When the population size is large, the average long-run cost per unit of time per system is obtained by taking the expectation of $c(\tau|Z\alpha,\beta,\rho)$:

$$E[c(\tau|Z\alpha,\beta,\rho)]) = \frac{C_p + C_c \cdot E[M(\tau, 0|Z\alpha,\beta,\rho)]}{\tau}. \tag{4.30}$$

Eq.(4.30) shows that the frailty has a direct impact on the renewal-type function. A thorough study of how $M(t, v|Z\alpha,\beta,\rho)$ is influenced by Z is part of future work but is beyond the scope of the current report. Since there exists no explicit formula for $M(t, v|\alpha,\beta,\rho)$, Monte Carlo simulation is used to determine the optimal replacement period τ^*. The following example demonstrates how a homogeneous population differs from a heterogeneous one in maintenance cost and optimal PM period.

Example 4.1.2 *A heterogeneous population is composed of $M = 25000$ independent ARA_∞ sequences with $\alpha = 1, \beta = 3, \rho = 0.6$. The heterogeneity is governed by the variance of α, $\sigma^2(\alpha)$, ranging in $\{0, 0.1, 0.2, 0.3\}$. α follows a gamma distribution with mean 1. The repair costs are set as $C_p = 1, C_c = 5$. In Figure 4.5, the long-run repair cost rate per system for a homogeneous population (when $\sigma^2(\alpha) = 0$) is shown by the blue line with a circle marker. When frailty is introduced, it decreases monotonically with the amplitude of heterogeneity (red, yellow, and purple curves). However, in this example, it is not clear how the optimal PM interval, τ^*, changes with $\sigma^2(\alpha)$.*

The model parameters are seldom known in real life: they have to be estimated from

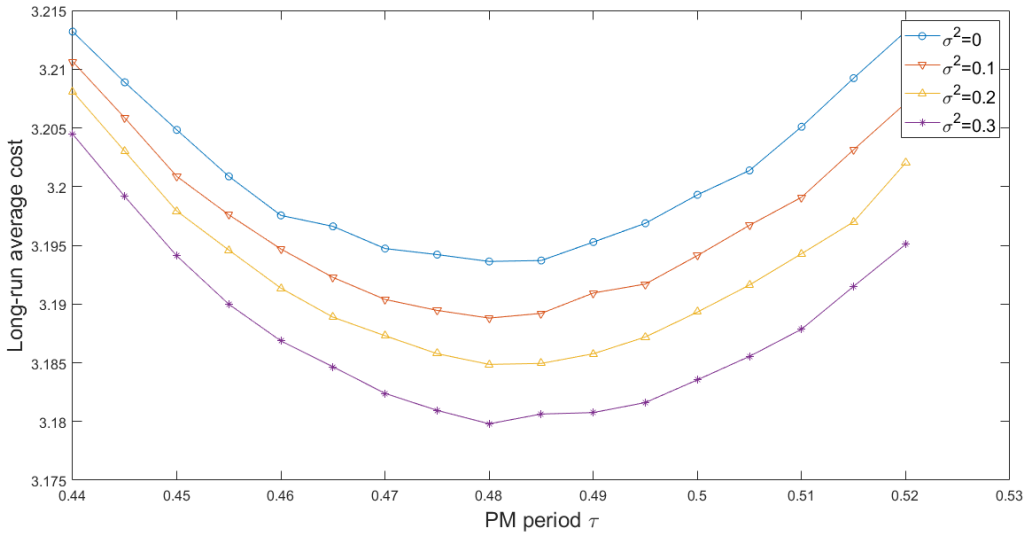


Figure 4.5: Long-run repair cost rate per system.

existing failure/repair records. We have previously shown that the parameters estimated under the faulty assumption that the population is homogeneous are proved to be biased. When scheduling PM activities for a heterogeneous population, the biased estimates will lead to non-optimal PM intervals, as is shown in the following example. We are not advocating that an optimal PM interval be assigned individually for each item: this usually leads to cumbersome computation and is not easy to implement or administer. Instead, we argue that an optimal PM period, same for all items in the population, could be obtained by accounting for the heterogeneity.

Example 4.1.3 *Consider a fleet of* $M = 100$ *independent repairable systems under imperfect repair of type* ARA_∞*. The individuals have different baseline failure intensities governed by the pseudo scale parameter* $\alpha_j, j \in 1...M$*, which follows a gamma distribution with mean 1 and variance 0.2 (*$k = 5, \theta = 0.2$*). Other parameters are set as* $\beta = 3, \rho = 0.6, C_c = 5, C_p = 1$*. A total of* $N = 50$ *observations are recorded for each system (generated with Monte Carlo simulation). Therefore, the failure histories of the population consist of a matrix of dimension* 100×50*, based on which we will determine the optimal PM interval.*

Three circumstances are considered: A) the PM interval is determined using the true parameters: this is the ideal situation and serves as a reference; B) We ignore the heterogeneity, and assume that all systems in the ARA_∞ *population are identical (share the same parameters); C) We assume that the baseline failure intensities of the systems differ only in* α *(with the same* β *and* ρ*).*

Circumstance A: When the true parameters α_j, β, ρ *are known,* $\bar{c}(\tau|\alpha_j, \beta, \rho) = 1/100 \sum_{j=1}^{100} c(\tau|\alpha_j, \beta, \rho)$ *is drawn by the orange line with cross marker in Figure 4.6.* $c(\tau|\alpha_j, \beta, \rho)$ *is defined by Eq.(4.28), representing the long-run repair cost per unit of time for system* j *when the PM interval is* τ*. It is obtained via Monte Carlo simulation due to the intractability of the renewal-type function. The true optimal PM interval* $\tau^* \simeq 0.47$ *(the orange diamond in Figure 4.6), with* $\bar{c}(\tau^*|\alpha_j, \beta, \rho) = 3.207$*.*

Circumstance B: Under the incorrect assumption that all items are identical, the triple $(\hat{\alpha}, \hat{\beta}, \hat{\rho})$ *is estimated as* $(1.4717, 2.2932, 0.758)$*: β and ρ are under/over estimated, which is consistent with the previous findings. For such a homogeneous ARA_∞ population, $c(\tau|\hat{\alpha}, \hat{\beta}, \hat{\rho})$ is shown by the blue line with cycle marker in Figure 4.6. $\hat{\tau} \simeq 0.405$ is the optimal PM period.*

Circumstance C: Taking the heterogeneity into consideration, the ARA_∞ parameters are estimated as $\tilde{\beta} = 3.0891, \tilde{\rho} = 0.5960$ using the inference method given in section 5.2.1. The pseudo parameters $\tilde{\alpha}_j, j \in 1...M$ are estimated individually for each item with Eq.(4.20). For such a heterogeneous ARA_∞ population, $\bar{c}(\tau|\tilde{\alpha}_j, \tilde{\beta}, \tilde{\rho})$ is drawn by the red line with triangle marker in Figure 4.6, and the optimal PM period $\tilde{\tau} \simeq 0.485$.

The advantage of considering the heterogeneity among the individual systems when scheduling PM activities is therefore highlighted by the fact that, red line is much closer to the orange line compared to the blue line. This is further verified by comparing $\tilde{\tau}$ and $\hat{\tau}$: $\bar{c}(\tilde{\tau}|\alpha_j, \beta, \rho) = 3.212$, whereas $\bar{c}(\hat{\tau}|\alpha_j, \beta, \rho) = 3.285$. The maintenance cost is therefore reduced when implementing PM with period $\tilde{\tau}$ on the ARA_∞ population governed by α_j, β, ρ.

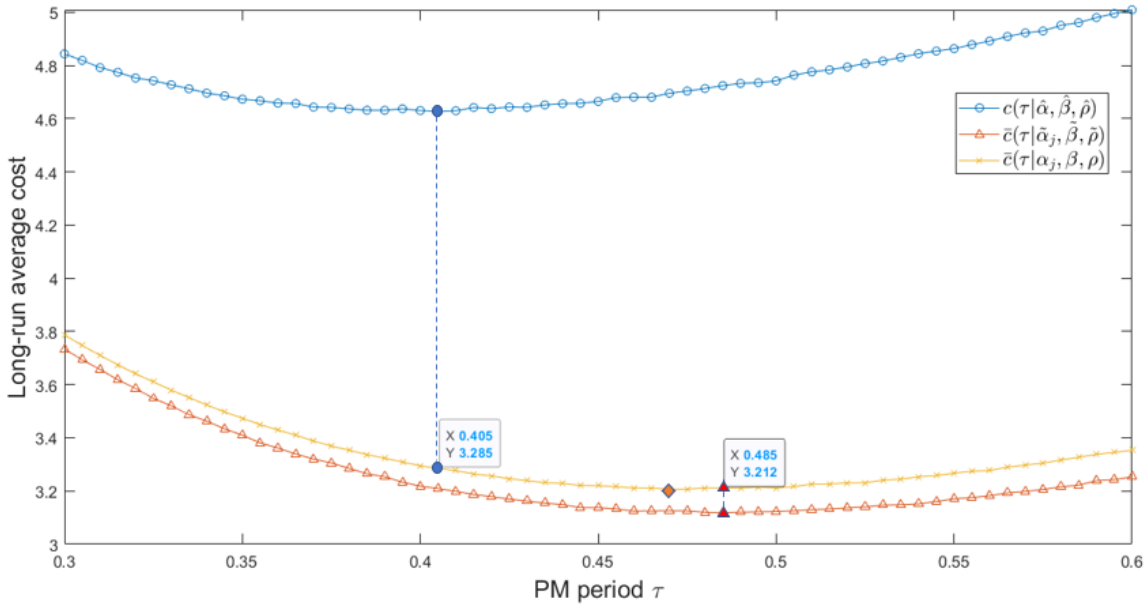


Figure 4.6: Long-run repair cost rate per system computed with the correctly specified model (heterogeneity taken into account, red line with triangle marker) and that computed with the erroneous model (heterogeneity ignored, blue line with cycle marker). The former is obviously closer to the real cost rate (orange line with cross marker) computed with true parameters.

It should be noticed that the significant reduction of long-run repair cost in Example 4.1.3 is related to the large size (100×50) of the simulated dataset: the more data we have, the more accurately we can estimate the parameters and the more the red line converges to the orange line. When the data size is limited, the red line's position may have a large variation (depending on the dataset itself), and a more cost-effective PM interval is not necessarily achieved by accounting for the heterogeneity, due to the potential bias of the estimates.

Considering the heterogeneity, if we further assign a specific PM interval for each system, the average long-run repair cost rate for the population depicted in Example 4.1.3 is around 3.088, which is the mean value of the minimums of 100 different cost curves. Although administering such a PM policy is generally troublesome, the maintenance cost is reduced significantly compared to the situation where all systems are preventively maintained with the same period. In Bane NOR, a compromised approach which consists in classifying assets into different groups (subpopulations) before planning the maintenance activities accordingly, is frequently used. This allows a balance between the practicability of the maintenance strategy and the maintenance expenditure.

4.2 Brown Proschan

BP model is another widely applied stable process used to characterize imperfect maintenance. When a fleet of repairable systems undergoes repair of type BP, i.e., the repair is perfect with a probability p and minimal with a probability $1-p$, the heterogeneity is shown to play an important role in the population's mean lifetime and reliability. In this section, we address the heterogeneous BP population from a mathematical point of view, and the issue of statistical inference is not pursued.

Imperfect repair models commonly have three parameters: two of them define the baseline failure rate, i.e., power law function $\lambda(t) = \alpha\beta t^{\beta-1}$, and a repair efficiency parameter ρ or perfect repair probability p that measures how efficient the maintenance is. For the ARA_∞ process, we have explicitly considered the frailty model where the heterogeneity is characterized by the proportional hazard rate of the individual systems. If the systems are working in different environments, the aging speed, often described by the Weibull shape parameter β, may present a heterogeneity because some harsh environments are more likely to accelerate the wear. If workers of different skills or experience maintain the systems, then it is also possible that identical systems will undergo repairs of different efficiency, and will present distinct mean lifetime and requires different maintenance resources.

In this section, for the BP model, we consider three types of heterogeneity: frailty (proportional failure rate), heterogeneous perfect repair probability, and heterogeneous shape parameter when the baseline distribution is Weibull. Particularly, we focus on the influence of heterogeneity on the population mean lifetime.

4.2.1 Frailty in BP

Let Z be the frailty variable with expectation 1. Similarly to ARA_∞ process, the failure intensity becomes therefore proportional, and the mean lifetime of a single BP process in its stable state given Z is:

$$\mu_Z = E[X_\infty|Z] = p\int_0^\infty e^{-pZ\Lambda(x)}dx, \tag{4.31}$$

which is a convex function of Z. Following Jensen's inequality, the expectation of μ_Z is larger than μ, the mean lifetime without heterogeneity:

$$E[\mu_Z] \geq \mu, \;\; \mu = p\int_0^\infty e^{-p\Lambda(x)}dx. \tag{4.32}$$

Same as in an ARA_∞ population, the heterogeneity will enlarge the mean lifetime of the assets in the population. We can then discuss the existence of $E[\mu_Z]$ by specifying the distribution of Z, denoted by f_Z, and the baseline failure rate $\Lambda(t)$. For example, consider the power law intensity:

$$\Lambda(t) = (\frac{t}{\eta})^\beta = \alpha t^\beta, \tag{4.33}$$

then the mean lifetime of a BP process when it enters stable state given the frailty Z is:

$$\mu_Z = p\int_0^\infty e^{-pZ\alpha t^\beta}dx = \mu Z^{-1/\beta}, \tag{4.34}$$

and the mean lifetime of the heterogeneous population is:

$$E[\mu_Z] = \mu E[Z^{-1/\beta}]. \tag{4.35}$$

This results holds for ARA_∞ population as shown in Eq.(4.1). Consequently, if Z is gamma distributed as in Eq.(4.2) with parameters k and θ, then the population mean lifetime $E[\mu_Z]$ is finite only if $\beta > \frac{1}{k}$, or, equivalently, $c_v(Z) < \sqrt{\beta}$. If the variation of Z is too large, the mean lifetime of the population tends to infinity, as shown in Example 4.2.2.

It is also important to look into the population's reliability function. Let X_∞^Z be the steady-state population cycle duration. X_∞^Z does not represent any individual steady-state cycle duration but could be regarded as the duration of a cycle drawn from the BP population when all the members have entered the stable regime. Similarly, let A_∞^Z be the steady-state population's VA, which is the VA after a repair of an item randomly drawn from the population that has entered the steady-state. The survival functions of A_∞^Z and X_∞^Z, when the frailty Z is gamma distributed with parameter k and $\theta = 1/k$, are respectively given by

$$R_{A_\infty^Z}(x|k) = (1-p)(1+\frac{p\Lambda(x)}{k})^{-k}, \tag{4.36}$$

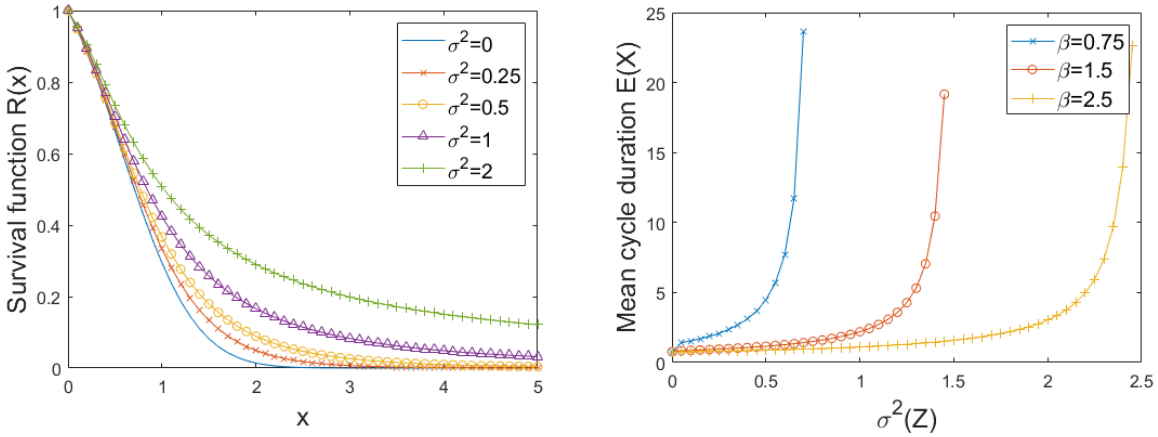
and

$$R_{X_\infty^Z}(x|k) = p\int_0^\infty \lambda(x+v)(1+\frac{\Lambda(x+v)-(1-p)\Lambda(x)}{k})^{-k-1}dv. \tag{4.37}$$

Example 4.2.1 *Consider the gamma-distributed frailty variable Z and the power law intensity $\lambda(x) = x^2$. The perfect repair probability p equals to 0.75. Figure 4.7a shows that $R_{X_\infty^Z}(x|k)$ is increasing in $\sigma^2(Z)$: when $\sigma^2(Z)$ rises from 0 to 2, the tail of the survival function becomes heavier and heavier, resulting in eventually an infinite value of the integral.*

Example 4.2.2 *Let Z be gamma distributed and $\beta \in \{0.75, 1.5, 2.5\}$, $p = 0.75$. Figure*

4.7b shows that the mean population lifetime $E(X_\infty^Z|k)$ *is increasing in* $\sigma^2(Z)$*, and that* $E(X_\infty^Z|k)$ *rises to infinity once* $\sigma^2(Z) > \beta$*.*



(a) $R_{X_\infty^Z}(x|k)$ increases with $\sigma^2(Z)$ (b) $E(X_\infty^Z|k)$ increases with $\sigma^2(Z)$

Figure 4.7: Examples of the survival functions (left) and mean cycle durations (right) of a heterogeneous BP population. The larger the variation of Z, the larger the survival function of X_∞^Z and the larger its expected value.

We can now investigate the inverse of μ_Z and its expectation. Previously, we have shown for ARA_∞ population that $E[1/\mu_Z]$ is vital for the calculation of the long-run repair cost for specific maintenance efficiency. For the Brown Proschan process, if the perfect repair probability p is a decision variable, a similar optimization framework could be established by assuming that the cost of a repair is an increasing function of p. In this way, the long-run average repair cost is proportional to $E[1/\mu_Z]$, as shown in Eq.(4.27).

Let Z be gamma distributed with parameter k and θ and let $E(Z) = k\theta = 1$. The survival function of $1/\mu_Z$ is given by

$$P(\frac{1}{\mu_Z} > t) = P(Z > (\mu t)^\beta) = \frac{\Gamma(k, \frac{(\mu t)^\beta}{\theta})}{\Gamma(k)}, \tag{4.38}$$

where $\Gamma(s, x)$ represents the upper incomplete Gamma function:

$$\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt. \tag{4.39}$$

Integrating the survival function, the inverse of the mean cycle durations has the following expectation:

$$E[\frac{1}{\mu_Z}] = \frac{1}{\mu}\theta^{1/\beta}\frac{\Gamma(k+1/\beta)}{\Gamma(k)} = \frac{1}{\mu}k^{-1/\beta}\frac{\Gamma(k+1/\beta)}{\Gamma(k)}. \tag{4.40}$$

Once again, the same result can be found in ARA_∞ processes, as shown in Eq.(4.6). This is because when the baseline failure rate is power law, the mean cycle duration of a BP or ARA_∞ process entering its stable state is proportional to its scale parameter η. For other baseline distribution, e.g., gamma or log-normal distribution, Eq.(4.40) does not necessarily hold.

4.2.2 Heterogeneous perfect repair probability

When the maintenance crew members are of different skill levels, the repair effectiveness could be dissimilar. Most likely, we would face the situation where the assets could be divided into subpopulations, each maintained by one maintenance personnel. In the case of BP imperfect repair models, the most skilled technician is more likely to conduct a perfect repair because they rarely make mistakes, whereas the less experienced one may have a higher probability of carrying out a minimal repair. Let us first look at a discrete mixture.

4.2.2.1 Discrete mixture

Consider a population composed of identical systems, governed by the baseline failure rate $\Lambda(t)$, undergoing imperfect repair of type BP. It is divided into two subpopulations, the first maintained by a skilled technician who is able to repair perfectly the systems with probability p_1 or minimally restore the systems with probability $1-p_1$; the second subpopulation is maintained by a less experienced worker, with a perfect repair probability p_2, and minimal repair probability $1-p_2$. $p_1 > p_2$.

The mean lifetimes when the system enters its stable state for the first subpopulation, μ_1, and that for the second subpopulation, μ_2, are respectively given by

$$\mu_1 = p_1 \int_0^\infty e^{-p_1\Lambda(x)}dx, \quad \mu_2 = p_2 \int_0^\infty e^{-p_2\Lambda(x)}dx. \tag{4.41}$$

Clearly, $\mu_1 > \mu_2$ if the baseline failure rate is increasing: for aging systems, the lifetime after a renewal is larger than that after a minimal repair. For the mixed population, the mean lifetime is given by its harmonic mean:

$$\mu = \frac{2}{\frac{1}{\mu_1} + \frac{1}{\mu_2}} = \frac{2\mu_1\mu_2}{\mu_1 + \mu_2}. \tag{4.42}$$

This being, an interesting question arises: is it better to employ two technicians having heterogeneous repair skills, i.e., p_1 and p_2, or to train the maintenance crew so that they have the same skill, i.e., an averagely satisfying perfect repair probability $\bar{p} = \frac{(p_1+p_2)}{2}$? Currently, the criteria are simple: we want to maximize the mean population lifetime, or in other words, we would like to minimize the total number of failures of the systems in the long term.

Assume that both of the technicians have a perfect repair probability, $\bar{p}$. Then the population is homogeneous: they share the same baseline failure rate and undergo homogeneous repair actions. Thus, the population mean lifetime is simply the mean cycle duration of a single BP process:

$$\nu = \bar{p} \int_0^\infty e^{-\bar{p}\lambda(x)}. \tag{4.43}$$

When the baseline failure intensity is described by an increasing power law function, i.e., Eq.(4.33) with $\beta > 1$, the population mean lifetime when the repairs are heterogeneous,

μ, is smaller than ν :

$$\mu = \frac{2\mu_1\mu_2}{\mu_1 + \mu_2} \leq \sqrt{\mu_1\mu_2} \leq \nu. \tag{4.44}$$

The first inequality results from harmonic mean being smaller than the geometric mean, and the second inequality comes from the power law failure rate:

$$\mu_1\mu_2 = (p_1p_2)^{1-1/\beta}(\eta\Gamma(1+\frac{1}{\beta}))^2, \tag{4.45}$$

and

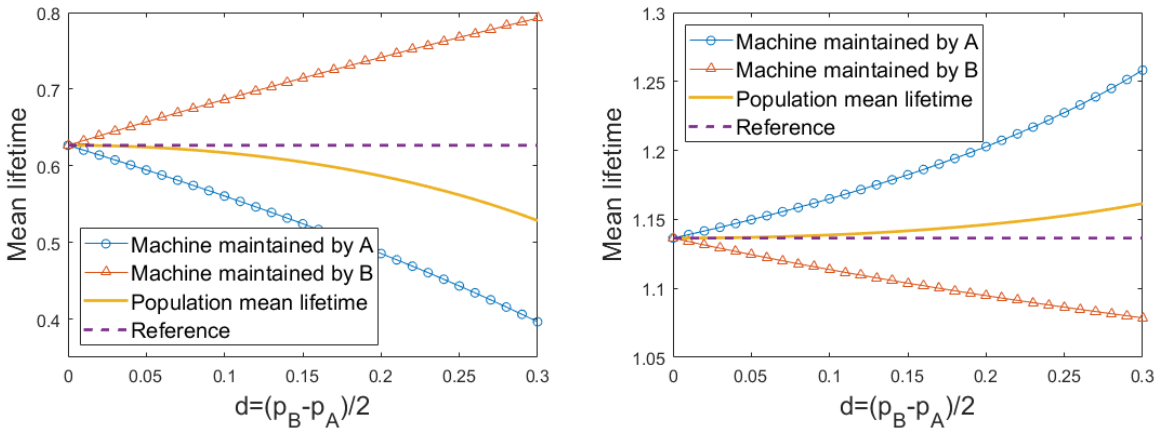
$$\nu^2 = (\frac{p_1+p_2}{2})^{2-2/\beta}(\eta\Gamma(1+\frac{1}{\beta}))^2. \tag{4.46}$$

Obviously, $(\frac{p_1+p_2}{2})^2 \geq p_1p_2$, and therefore $\nu \geq \sqrt{\mu_1\mu_2}$. This means in order to minimize the number of system failures over a long period of time, it is better to employ some homogeneous maintenance crew who have similarly adequate skill/experience, rather than a mixture of some highly skillful personnel and inexperienced rookies. On the contrary, if the baseline failure rate $\lambda(t)$ is decreasing, i.e., systems are getting younger as time goes by, then the inequality in Eq.(4.44) is inversed, which means it would be better to employ rookies who can generally ensure a minimal repair to minimize the long-run number of failures. This is demonstrated by the following example.

Example 4.2.3 *Consider two independent and identical machines maintained by technician A and B. The repair process is described by BP, where the perfect repair probabilities of technician A and B are denoted respectively p_A and p_B with $p_A \leq p_B$. Assume that the average perfect repair probability is 0.5. Let $d = (p_B - p_A)/2$ be a measure of the heterogeneity of the perfect repair probability. Figure 4.8a depicts how the mean lifetimes of the two machines, as well as the mean lifetime of the population, change with d when the aging parameter $\beta = 2$: the dashed line represents the mean population lifetime when $p_A = p_B = 0.5$. When the assets are aging, the machine maintained by technician B (A) has a longer (shorter) mean lifetime but overall, the mean cycle duration of the population is decreasing in d; conversely, the mean lifetime of the population is increasing in d when the systems are getting younger ($\beta = 0.9$) as shown in Figure 4.8b. The results are consistent with Eq.(4.44).*

4.2.2.2 Continuous mixture

To complete this section, we shall discuss a continuous mixture. This happens when our maintenance crew has an instable performance: in average, the personnel has a probability p to restore the system to the as-good-as-new state; but sometimes he or she can do better and sometimes much worse, depending, for example, on the mental state or some external factors. This being, the performance of the maintenance crew, i.e., the perfect repair probability p, can be regarded as a continuous random variable defined on $[0, 1]$, which suggests implicitly that better-than-new and worse-than-old repairs are excluded from the current discussion.



(a) $\beta = 2$ (b) $\beta = 0.9$

Figure 4.8: Variation of the mean lifetime. The blue curve and red curve represent respectively the mean lifetime of machines maintained by A and B. The population mean lifetime is their harmonic mean, drawn by the orange line. The purple dashed line represents the homogeneous situation where the two technicians share the same perfect repair probability.

Let f_p be the probability density/mass function of p and $\bar{p} = E[p]$ be the mean probability of perfect repair. For now, we do not specify f_p. Let X_∞^p be the steady-state population cycle duration. It should be noted that we adopt the superscript 'p' rather than 'Z' because the source of the heterogeneity is no longer a frailty variable. This being, the expected value of X_∞^p is given by:

$$E[X_\infty^p | f_p] = \int_0^1 \int_0^\infty s e^{-s\Lambda(x)} f_p(s) dx ds. \tag{4.47}$$

From a practical point of view, it is useful to define an 'average' mean lifetime, which represents the mean lifetime of a BP population that undergoes homogeneous repairs with perfect repair probability $\bar{p} = E[p]$:

$$E[X_\infty | \bar{p}] = \bar{p} \int_0^\infty e^{-\bar{p}\Lambda(x)} dx. \tag{4.48}$$

The relation between $E[X_\infty^p | f_p]$, and its 'averaged' version $E[X_\infty | \bar{p}]$, is given below.

Proposition 3 *Consider the power law baseline failure intensity: $\lambda(x) = \alpha\beta x^{\beta-1}$. Then,*

$$E[X_\infty^p | f_p, \beta] \leq E[X_\infty | \bar{p}, \beta] \quad \textit{if} \quad \beta \geq 1, \tag{4.49}$$

and

$$E[X_\infty^p | f_p, \beta] \geq E[X_\infty | \bar{p}, \beta] \quad \textit{if} \quad \beta \leq 1. \tag{4.50}$$

proof 7 *Let $h(p) = p \int_0^\infty e^{-p\alpha x^\beta} dx$. When $\beta > 1$, αx^β is a convex function of x, making*

$h(p)$ a concave, but increasing function of p. Applying Jensen's inequality: $E[h(p)] \leq h[E(p)]$ leads to $E[X_\infty^p|f_p,\beta] \leq E[X_\infty|\bar{p},\beta]$. Conversely, when $\beta < 1$, αx^β is concave and $h(p)$ is convex, resulting in $E[X_\infty^p|f_p,\beta] \geq E[X_\infty|\bar{p},\beta]$.

This result can be interpreted as follows: consider two fleets of identical repairable systems under imperfect repair of type BP. The first fleet is repaired by workers having homogeneous skills: each item is maintained to perfect state with probability p. Technicians of different expertise maintain the assets in the second fleet: some are more likely to repair the asset to a perfect state than others, but on average, the probability of perfect repair is also p. This being, if the assets under consideration are wearing out ($\beta > 1$), the average lifetime of the assets in the second fleet is smaller than that of the assets in the first fleet. This is consistent with what has been observed for the discrete mixture.

The Beta distribution is usually utilized to model the behavior of random variables limited to intervals of finite length, particularly the proportion or probability outcomes. Assume now that the perfect repair probability p in BP models follows a Beta distribution:

$$f_p(s) = \frac{s^{a-1}(1-s)^{b-1}}{B(a,b)}, a > 0, b > 0, \tag{4.51}$$

where $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the Beta function. Then the expected value and survival function of the asymptotic cycle duration of the population are given by:

$$\begin{aligned} E[X_\infty^p|f_p] &= \int_0^1 \int_0^\infty s e^{-s\Lambda(x)} \frac{s^{a-1}(1-s)^{b-1}}{B(a,b)} dx ds \\ &= \frac{a}{a+b} \int_0^\infty \Phi(a+1; a+b+1; -\Lambda(x)) dx, \end{aligned} \tag{4.52}$$

and

$$R_{X_\infty^p}(x|f_p) = \frac{a}{a+b} \int_0^\infty \lambda(x+v) e^{-\Lambda(x+v)+\Lambda(v)} \cdot \Phi(a+1; a+b+1; -\Lambda(v)) dv, \tag{4.53}$$

where $\Phi(b;c;z) = \sum_{n=0}^\infty \frac{B(b+n,c-b)}{B(b,c-b)} \frac{z^n}{n!}$ is the extended confluent hyper-geometric function.

Unlike heterogeneous failure intensity, the variance of p will not lead to infinite mean cycle duration of the population when the asset is aging: $E[X_\infty^p]$ reaches its maximum (the mean lifetime of a renewal process) when p constantly equals to 1; However, when $\beta < 1$, the existence of $E[X_\infty^p]$ depends on f_p and on λ.

Example 4.2.4 *Let $\lambda(x) = x^2$. The perfect repair probability p follows a Beta distribution with mean $E[p] = 0.5$ and variance $\sigma^2(p) = [0.05, 0.1, 0.15, 0.2]$. The pdf of p are drawn in Figure 4.9. The corresponding survival functions of the population lifetime are shown in Figure 4.10. It is observed that $\forall x > 0$, the larger $\sigma^2(p)$, the smaller $R_{X_\infty^p}(x|f_p)$ would be.*

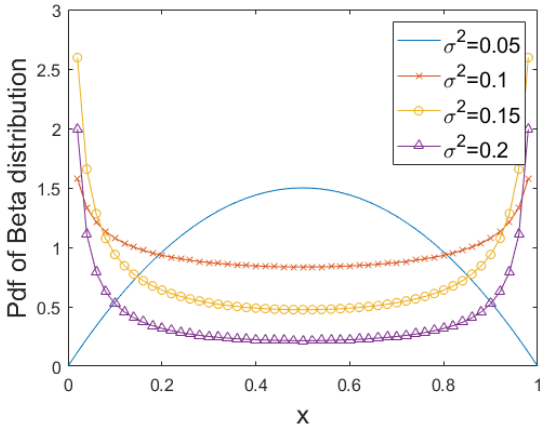


Figure 4.9: Pdf of p for different $\sigma^2(p)$. $E[p] = 0.5$.

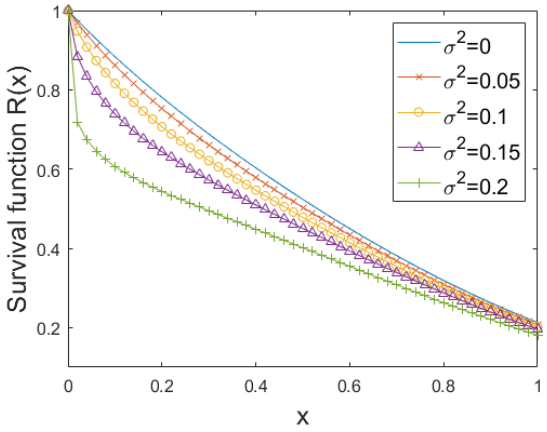


Figure 4.10: $R_{X_\infty^p}(x|f_p)$ as a function of $\sigma^2(p)$.

4.2.3 Heterogeneous aging parameter

Assume that the first intervals of all systems in the population are Weibull distributed. This corresponds to the power law failure intensity: $\lambda(x) = \alpha\beta x^{\beta-1}$. Under different working conditions, the assets may have different wearing velocities, resulting in a non-homogeneous aging parameter β. We assume that there is no frailty variable and that repair effectiveness, p, is homogeneous. We investigate in this section how the heterogeneous β influences the population mean lifetime.

Let f_β be the probability density/mass function of β and $\bar{\beta} = \int_0^\infty s f_\beta(s) ds$ be its expected value. Let X_∞^β be the steady-state population cycle duration. It should be noted that we adopt the superscript 'β' rather than 'Z' or 'p' because the source of the heterogeneity is neither a frailty variable nor the repair effectiveness.

The convexity of $E[X_\infty]$ as a function of β, depends on β. Therefore, no direct result can be drawn regarding the relationship between $E[X_\infty|\bar{\beta}]$, the "average" mean cycle duration and $E[X_\infty^\beta|f_\beta]$, the true mean lifetime of the heterogeneous population. This is shown in the following example.

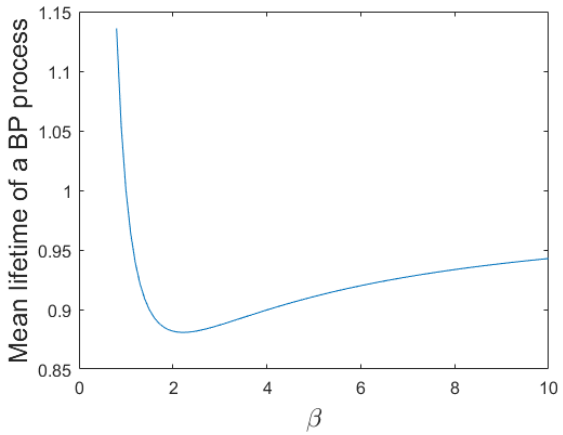


Figure 4.11: $E[X_\infty|\beta]$ as a function of β.

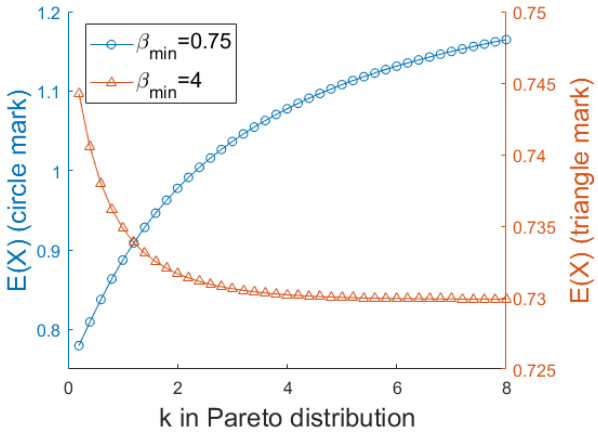


Figure 4.12: Monotonicity of $E[X_\infty|f_\beta]$.

Example 4.2.5 *Let $\lambda(x) = \beta x^{\beta-1}$ be the failure intensity and $g(\beta) = E[X_\infty|\beta] = p\int_0^\infty e^{-px^\beta}dx$ be a function of β. Figure 4.11 shows that $E[X_\infty|\beta]$ is first a convex then a concave function of β. The inflection point β^* (where the second derivative changes sign) when $p = 0.99$ is around 3.45: $E[X_\infty|\beta]$ is a convex function before 3.45 and a concave one after.*

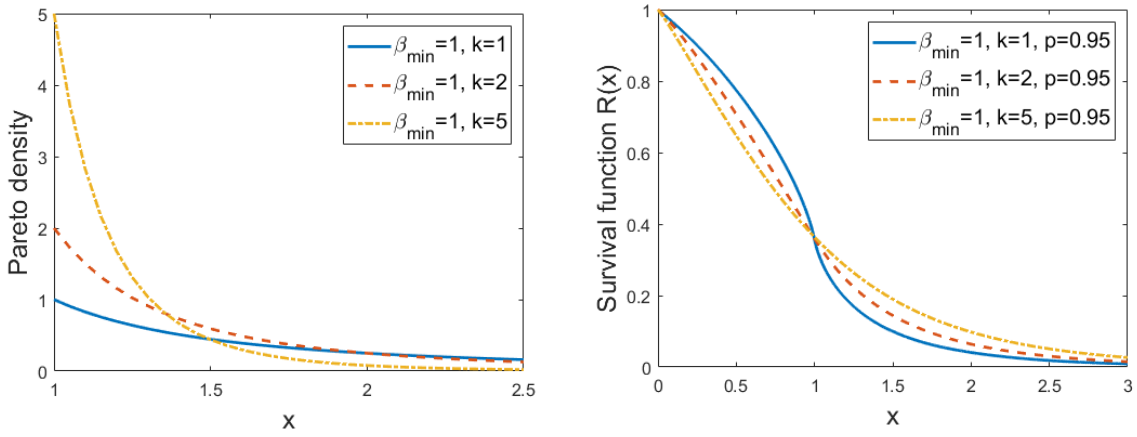
Consider the situation where the diverse aging parameters of the n systems in a population, $\beta_1, \beta_2 ... \beta_n$, are deterministic and known. $\bar{\beta} = \sum_{i=1}^n \beta_n/n$ is the mean aging rate. The mean lifetime of the population, $E[X_\infty^\beta|\beta_1, \beta_2 ... \beta_n]$ is the harmonic mean of the mean cycle durations of each individual system. When $max(\beta_i) < \beta^*$, all the shape parameters belong to the domain on which $E[X_\infty|\beta]$ is convex, resulting in $E[X_\infty^\beta|\beta_1, \beta_2 ... \beta_n] \geq E[X_\infty|\bar{\beta}]$. Conversely, if $min(\beta_i) > \beta^*$, the concavity leads to $E[X_\infty^\beta|\beta_1, \beta_2 ... \beta_n] \leq E[X_\infty|\bar{\beta}]$.

Now, let us consider a specific distribution for β. If a reliability engineer says: "We know the machines are aging, and that 80% of the aging rates are smaller than 2", then this may be referred to as the "80-20 rule", or the Pareto principle, which may further inspire the utilization of the Pareto distribution to describe the empirical dispersion of the Weibull shape parameters.

Let f_β be the Pdf of a Pareto law with scale parameter $\beta_{min} > 0$ and shape parameter $k > 0$:

$$f_\beta(s) = \frac{k\beta_{min}^k}{s^{k+1}}, \quad s \geq \beta_{min}. \tag{4.54}$$

β_{min} represents the lower bound of β and k decides the shape of the density: the larger the k, the more β cluster around β_{min}. Some Pdf of the Pareto distribution with $\beta_{min} = 1, k = 1, 2, 5$ are drawn in Figure 4.13a. For a Pareto distributed random variable, the mean exists only if $k > 1$, and that the variance exists only if $k > 2$.



(a) Pareto distribution

(b) $R_{X_\infty^\beta}(x|f_\beta)$

Figure 4.13: Pdf of Pareto distribution (left) and survival function of X_∞^β (right) when β follows a Pareto distribution. Parameters are set as $\beta_{min} = 1, k = 1, 2, 5$.

The population mean lifetime $E[X_\infty^\beta|f_\beta]$ is given by:

$$E[X_\infty^\beta|f_\beta] = \int_{\beta_{min}}^\infty \int_0^\infty pe^{-p(x/\eta)^s} \frac{k\beta_{min}^k}{s^{k+1}} dxds = \eta pk\beta_{min}^k \int_0^{\frac{1}{\beta_{min}}} p^{-t}\Gamma(1+t)t^{k-1}dt. \tag{4.55}$$

The monotonicity of $E[X_\infty|f_\beta]$ is relatively complex. For illustrative purpose, let's consider the special case where $\beta_{min} = 1$: the systems are simply wearing out. Then, the derivatives of $E[X_\infty^\beta|f_\beta]$ with respect to η, p and k are given by:

$$\begin{cases} \dfrac{\partial E(X_\infty|f_\beta)}{\partial \eta} = pk\displaystyle\int_0^1 p^{-t}\Gamma(1+t)t^k dt, \\ \dfrac{\partial E(X_\infty|f_\beta)}{\partial p} = \eta k\displaystyle\int_0^1 p^{-t}\Gamma(1+t)t^{k-1}(1-t)dt, \\ \dfrac{\partial E(X_\infty|f_\beta)}{\partial k} = \eta p\displaystyle\int_0^1 p^{-t}\Gamma(1+t)t^{k-1}(1+klog(t))dt. \end{cases} \tag{4.56}$$

When $\beta_{min} = 1$, the partial derivatives are both positive, indicating that $E[X_\infty^\beta|f_\beta]$ is increasing in η, p and k. This is because 1) in a Weibull distribution, the mean lifetime is proportional to η 2) when the asset is aging, the larger the p, the more we repair and the longer the average lifetime and 3) The larger the k, the more the shape parameters cluster around β_{min} which is 1, so the BP processes behave more like a renewal process.

Example 4.2.6 *Let $p = 0.95$ be the perfect repair probability. Figure 4.12 shows that the monotonicity of $E[X_\infty|f_\beta]$ with respect to k depends on β_{min}: the red curve with triangle marks reveals that $E[X_\infty|f_\beta]$ is decreasing in k when $\beta_{min} = 4$, whereas the blue curve with circle marks suggests that $E[X_\infty|f_\beta]$ is increasing in k when $\beta_{min} = 0.75$.*

4.3 Geometric Process and ARA_1

Assume that the baseline failure intensity is described by the power law: $\lambda(t) = \beta/\eta(t/\eta)^{\beta-1}$. For a Weibull distributed random variable, the scale parameter η is proportional to the mean lifetime. Thus, for ARA_1 and geometric process, we investigate the situation where the scale parameter itself is heterogeneous, e.g., characterized by a gamma distribution. This is slightly different from the previous studies where the heterogeneity is imposed on the pseudo parameter α with $\alpha = \eta^{-\beta}$: we are no longer in the framework of frailty analysis where the hazard rates of individual systems are proportional.

Let the heterogeneity among the individuals be characterized by a gamma-distributed Weibull scale parameter η, with shape parameter k and scale parameter θ:

$$f_\eta(s|k,\theta) = \frac{1}{\Gamma(k)\theta^k}s^{k-1}e^{-\frac{s}{\theta}}, \tag{4.57}$$

with mean $k\theta = \overline{\eta}$. Thus, $\overline{\eta}$ is an "average" scale parameter to some extent. On the other hand, the degree of heterogeneity between the individuals can be modeled by the variance, $\sigma^2(\eta)$, which equals to $k\theta^2$. k and θ are therefore fully determined by $\overline{\eta}$ and $\sigma^2(\eta)$. In the following, $\overline{\eta}$ is set to 1. We reveal for ARA_1 and GP how $\sigma^2(\eta)$ influence the estimation bias for the aging parameter and maintenance efficiency when the heterogeneity is erroneously overlooked. The issue of the population mean lifetime is not pursued, since the ARA_1 and geometric process, unlike ARA_∞ or BP, are not stable: the cycles are stochastically decreasing.

4.3.1 Geometric process

First introduced in [79], GP is used to describe inter failure times with trends. Let F be the Cdf of the first interval X_1. If $X_i, i \in \mathbb{N}^*$ are independent and the distribution function of X_i is given by $F(a^{i-1}x)$ where a is a positive constant, then $\{X_i, i \in \mathbb{N}\}$ is called a geometric process (GP). Obviously, if $a > 1$, $\{X\}$ is stochastically decreasing and converge to zero with probability 1; if $0 < a < 1$, $\{X\}$ is stochastically increasing and converge to infinity with probability 1; for $a = 1$, GP becomes a renewal process.

Monte Carlo simulation is used to generate several large datasets, from which we shall try to estimate the parameters. The model is misspecified, i.e., the heterogeneity is wrongly overlooked, and the scale parameters of each system are assumed to be identical. A total of $M = 5000$ independent GP is generated. The scale parameters η_j of the j-th sequence, $j \in 1..., M$, are gamma-distributed. Let β be the common shape parameter, and a be the common multiplier. $n_j = N = 200$ is the length of the j-th sequence and $X_{j,i}$ is the i-th interval in j-th sequence. We investigate the cases where $a \in \{0.5, 1.5, 3.5\}$ and $\beta \in \{0.75, 1.5, 2.5\}$. This being, the likelihood function given the observation matrix is:

$$\mathcal{L}_{GP}(\eta, \beta, \rho|\mathcal{X}_M) = -\beta log(\eta) \sum_{j=1}^{M} n_j + \frac{1}{2}\beta log(a) \sum_{j=1}^{M} n_j^2 - n_j + log(\beta) \sum_{j=1}^{M} n_j + (\beta - 1) \sum_{j=1}^{M} \sum_{i=1}^{n_j} log(X_{j,i}) - \eta^{-\beta} \sum_{j=1}^{M} \sum_{i=1}^{n_j} (\frac{X_{j,i}}{a^{1-i}})^{\beta}. \quad (4.58)$$

Obviously, Eq.(4.58) does not account for the heterogeneity. Unlike ARA_∞ or BP, the intervals in a GP sequence are independent, and the distribution of i-th interval is not influenced by the age just after the $i-1$ th repair. As a result, the multiplier a in GP population is neither affected by the heterogeneity of η, nor the value of β (Figure 4.14). In fact, $\hat{a}$ is always unbiased. β, on the other hand, is consistently underestimated, although the amount of underestimation depends only on β and $\sigma^2(\eta)$ and is not affected by a.

4.3.2 ARA_1 model

The ARA_1 population is generated in the same way as geometric process and the likelihood function to maximize is given in Eq.(4.9) with $a_{j,i} = a_{j,i-1} + (1-\rho)X_{j,i}$. This is because the main difference between ARA_1 and ARA_∞ lies in the age reduction mechanism. The configurations studied are: $\rho \in \{0.25, 0.5, 0.75\}$ and $\beta \in \{0.75, 1.5\}$. ARA_1 is more sensitive to heterogeneities than ARA_∞ or BP, and the bias of estimation when the model is mis-specified depends manifestly on $\sigma^2(\eta)$. Therefore, we focus first on a small scale where $\sigma^2(\eta) = 10^{-3} : 10^{-3} : 10^{-2}$, before investigating the estimation bias with $\sigma^2(\eta) = 0.02 : 0.02 : 0.5$.

When $\beta = 0.75$, ρ is overestimated and β is underestimated. In Figure 4.15a, the estimation bias for ρ is plotted: The blue curve with cross marker represents $\hat{\rho}$ when $\rho = 0.25$. As $\sigma^2(\eta)$ increases, ρ is more and more overestimated. The same is observed for $\rho = 0.75$ and

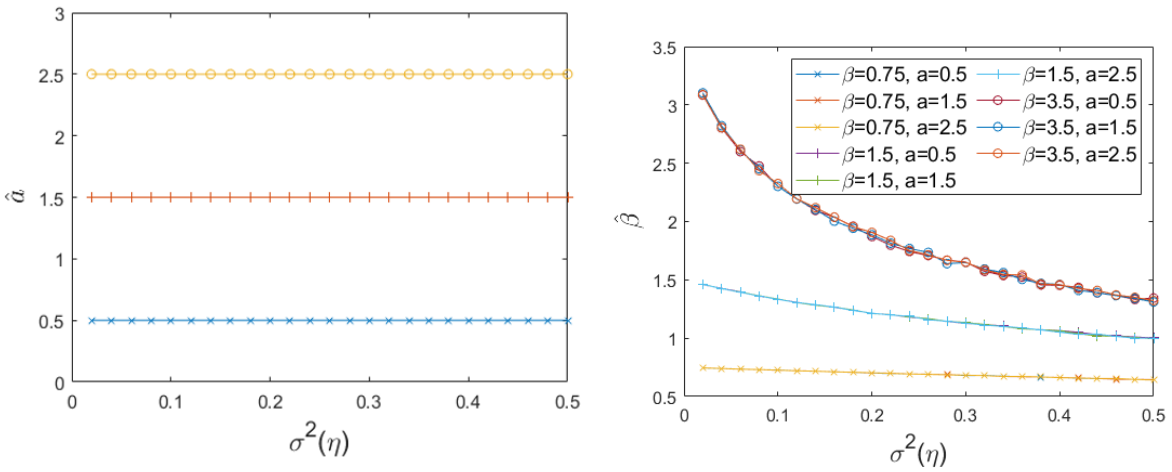
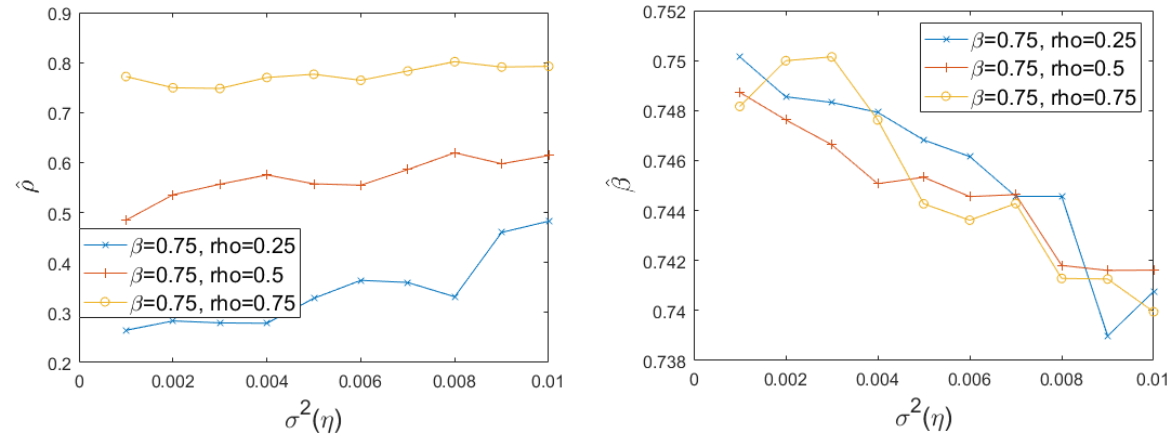


Figure 4.14: $\hat{a}$ (left) and $\hat{\beta}$ (right) in GP with $\sigma^2(\eta) = 0.02 : 0.02 : 0.5$. The estimation of multiplier a is drawn in the figure on the left: no matter what value β takes, $\hat{a}$ is unbiased. On the right side, the estimations of β are shown by different curves: circle marker for $\beta = 3.5$, plus marker for $\beta = 1.5$ and cross marker for $\beta = 0.75$. Clearly, regardless of a, $\hat{\beta}$ is decreasing in $\sigma^2(\eta)$, indicating that β is constantly underestimated, and that the amount of underestimation is increasing in the heterogeneity.

0.5, i.e., the yellow curve with the circle marker and the red curve with a plus marker. Figure 4.15b shows that β is underestimated and that the amount of underestimation seems to be independent of ρ, as suggested by the closeness of the curves.

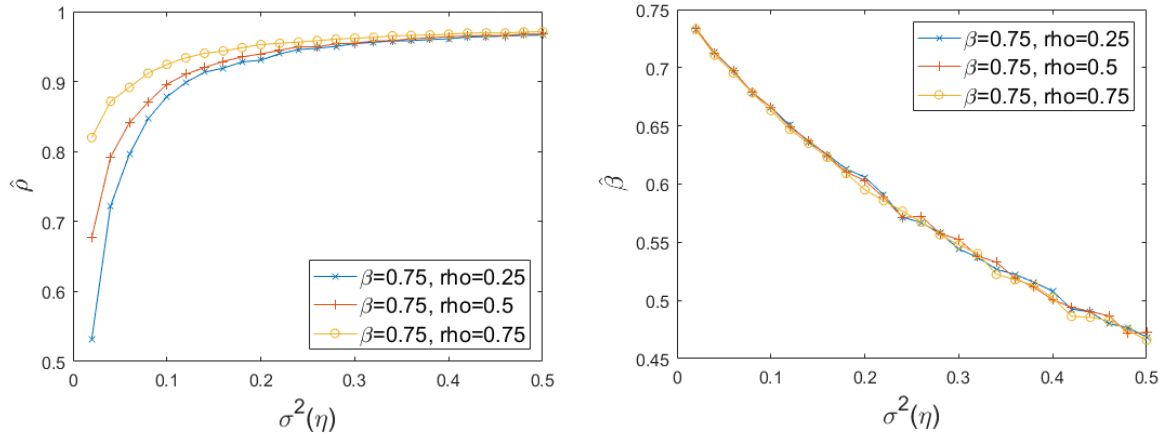


(a) $\hat{\rho}$ (b) $\hat{\beta}$

Figure 4.15: Parameter estimation with $\beta = 0.75$, $\sigma^2(\eta) = 10^{-3} : 10^{-3} : 10^{-2}$.

In Figure 4.16, the estimation bias for ρ and β is plotted on a large scale: the behaviors of the estimators are consistent to what has previously been observed: the Figure 4.16a shows that ρ is overestimated, whereas Figure 4.16b shows that β is underestimated.

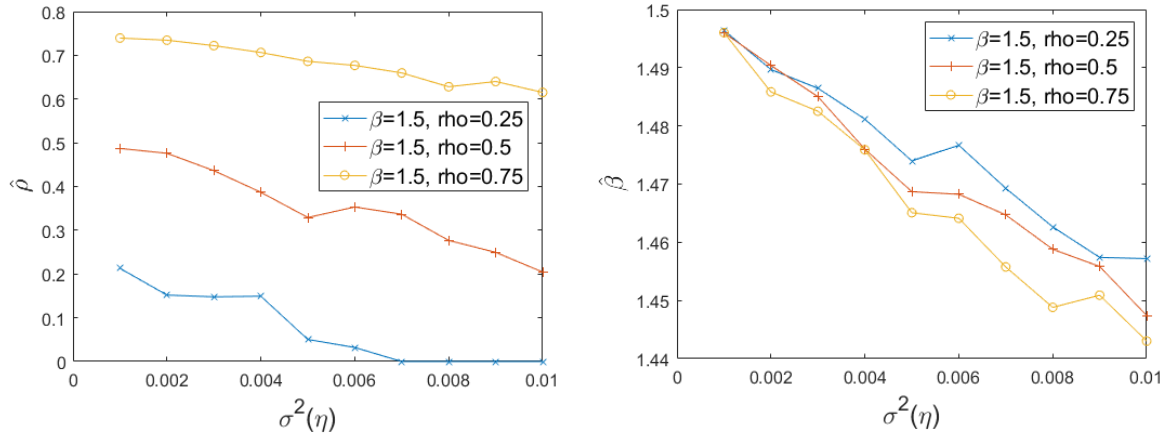
When $\beta = 1.5$, with a minor heterogeneity, ρ and β are underestimated and $\hat{\rho}$ quickly reaches to 0 (Figures 4.17). In contrary to the case $\beta = 0.75$, we can now observe, on a small scale, the underestimation of ρ in Figure 4.17a: the blue curve with cross marker representing $\hat{\rho}$ when $\rho = 0.25$ quickly drops to 0 as $\sigma^2(\eta)$ increases. The same trend is



(a) $\hat{\rho}$ (b) $\hat{\beta}$

Figure 4.16: Parameter estimation with $\beta = 0.75$ and $\sigma^2(\eta) = 0.02 : 0.02 : 0.5$.

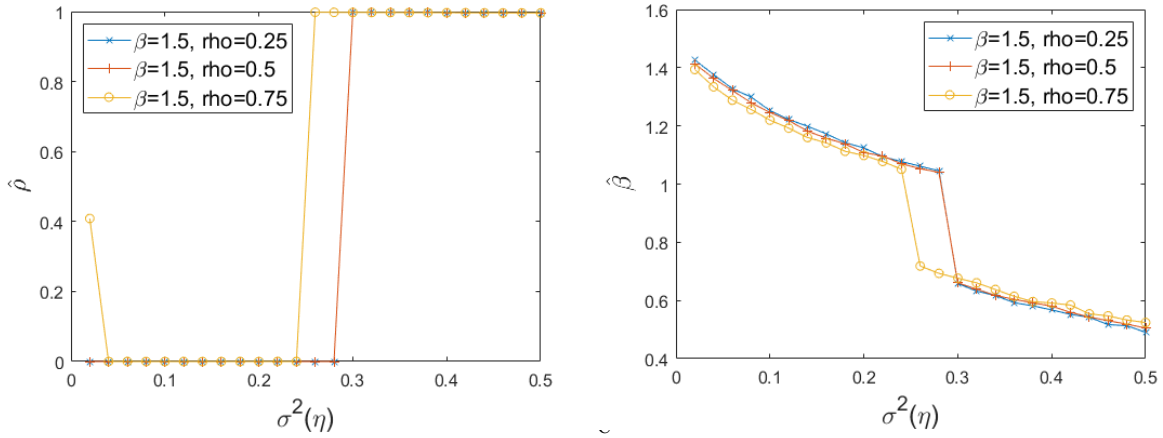
observed for $\rho = 0.75$ and 0.5, i.e., the yellow curve with the circle marker and the red curve with the plus marker. Nevertheless, Figure 4.17b shows that β is still underestimated, and the amount of underestimation is hardly influenced by ρ, as suggested by the closeness of the curves.



(a) $\hat{\rho}$ (b) $\hat{\beta}$

Figure 4.17: Parameter estimation with $\beta = 1.5$ and $\sigma^2(\eta) = 10^{-3} : 10^{-3} : 10^{-2}$.

If we zoom out and consider larger $\sigma^2(\eta)$, it is observed in Figure 4.18 that once the heterogeneity is large enough, $\hat{\rho}$ will increase dramatically to 1. The sudden rise leads to a swift drop of $\hat{\beta}$. In a larger scale, $\hat{\rho}$ remains 0 until the variance of η is too large: once $\sigma^2(\eta)$ reaches some values between 0.2 and 0.3, $\hat{\rho}$ jumps to 1, which obviously makes no sense because an ARA_1 process with a repair efficiency 1 is just a renewal process. The jump is also observed in ARA_∞ process but in the opposite direction. On the right side, β is always observed to be underestimated.



(a) $\hat{\rho}$ (b) $\hat{\beta}$

Figure 4.18: Parameter estimation with $\beta = 1.5$ and $\sigma^2(\eta) = 0.02 : 0.02 : 0.5$.

4.4 Conclusion

In this chapter, we have considered the case wherein heterogeneity among systems is combined with stable imperfect repair models. For ARA_∞ process, after investigating the influence of unspecified frailty on the population mean lifetime, we considered specifically the gamma distribution on the pseudo-scale parameter of the ARA_∞ process and derived the asymptotic properties, including the distributions of the population mean cycle duration and population VA and then presented the consequences for instances wherein the heterogeneity among systems was erroneously ignored. In particular, when the model was misspecified, the aging rate was underestimated, while the repair efficiency was overestimated. Furthermore, owing to the specialty of ARA_∞, that is, dependent intervals, the MLE established on the correct model was still inconsistent. Therefore, an alternative approach has been proposed, and its consistency was verified. Finally, the issue of maintenance optimization for the population that undergoes imperfect repair is addressed, and the benefits of considering the heterogeneity when scheduling preventive maintenance activities are demonstrated.

For the Brown Proschan model, we have investigated not only the frailty-originated heterogeneity but considered the heterogeneous maintenance effectiveness as well. It has been highlighted that in order to minimize the long-run average number of systems' failures, it is better to employ an 'averagely' skillful crew than a combination of highly skilled personnel and some rookies.

Finally, for two instable imperfect repair process, ARA_1 and geometric process, we show that estimating the model parameters without considering the hidden heterogeneity may lead to biased estimates. Specifically, the aging parameter, when the baseline distribution is Weibull, is always underestimated. This is an important fact that emphasizes the importance of accounting for the system's heterogeneity when analyzing the maintenance history; otherwise, with the underestimated aging parameter and an often overestimated repair efficiency, it is hard to optimize the maintenance plan.

The industrial applications of these results shall be demonstrated in Chapter 6, where a

case study on failure data collected from signals alongside a Norwegian railway station is carried out, illustrating the benefits of taking account of unobserved heterogeneity in the estimation of the aging speed and reliability of assets and in scheduling preventive maintenance activities.

Nevertheless, there are several limitations to the current study. First and foremost, no covariate is taken into consideration. In practice, the covariates explain the sources of the heterogeneity and allow us to measure/control it more proficiently, while in our study, the heterogeneity is regarded as an intrinsic property of the systems. This shall be addressed again in Chapter 6 when we analyze the failure history of the railway signaling systems operated by Bane NOR. Second, it would be interesting to investigate how different heterogeneities interact with each other, given that we have only studied *separately* the heterogeneous failure rate, heterogeneous repair effectiveness, and different aging velocity. We believe that these two directions merit further investigation.

Chapter 5

Approximation of the Superposition of Renewal Processes

The Superposition of Renewal Processes (SRP) consists of the observation of inter-arrival times based on multiple independent renewal processes. In imperfect maintenance analysis, SRP characterizes the repairable series system when maintenance consists of replacing the failed component while leaving the other components unchanged (minimally repaired). The components can be either physical or virtual structures as in a competing risk situation. After a repair, the system is often between the states as-good-as-new and as-bad-as-old since only one of the components has been renewed.

In practice, maintenance records are often incomplete, and the information on the identity of the failed components is not necessarily available. Consequently, the observations are commonly reduced to a pooled output [21], consisting of the failure times at a system level. Assessing the health of the system, such as its overall aging and maintenance efficiency, usually starts by estimating model parameters. When the number of components is known, and all components are identical, inference procedures have been carried out by Zhang et al. [118] directly on the pooled output using computational partitioning. Nevertheless, when the number of pooled events is limited, or when the components are different from each other, direct inference methods are difficult to implement, inspiring diverse approximation approaches.

Since the inter-occurrence times of an SRP tend to an equilibrium distribution, a natural simplification is to use a renewal process [69, 113, 105] to approximate the SRP. There exist multiple choices for the lifetime distribution of the approximating renewal process. Two of them are most commonly employed: the first is the exponential distribution, which forms a Homogeneous Poisson process (HPP) [113]; the second is the limiting distribution of intervals in the approximated SRP given in Eq.(2.72), which ensures that the approximating renewal process and the approximated SRP have the same expected lifetime. This is referred to in [105] as the stationary interval method (SIM).

However, the main drawback of approximating an SRP by a renewal process is the loss of dependency between the inter-occurrence times, in particular two successive inter-arrival times. In this chapter, we propose three novel approximation approaches of an SRP based on virtual age models and copula, before comparing them to existing models, e.g., HPP

and SIM. Since we are interested mainly in the situation where the SRP parameters are unknown, we shall present first the five approaches while emphasizing on the estimation procedures that derive the parameters of the approximating models from the observed pooled output of an SRP. Then, we evaluate the performance of the proposed methods, i.e., to see if they can correctly estimate the mean lifetime or capture the correlation between adjacent intervals or implement other methods to assess the "distance" between the SRP and its approximations.

5.1 Approximation methods

In this section, we introduce five models that will be later used to approximate an SRP and focus particularly on the statistical inference methods. But first, let us define the notations.

Consider a series system wherein corrective maintenance (CM), which consists of replacing the failed component while not maintaining the others, is carried out immediately after a failure. Periodic preventive maintenance (PM) is performed at $\Delta, 2\Delta, 3\Delta....$, and $\Delta = \infty$ if no PM is implemented. It is also assumed that at the beginning of the observations, the system is in its *stationary (equilibrium) regime*, which implies that it is not as-good-as-new at $t = 0$. The failure times are denoted by $\{T_i\}_{i\geq 1}$, with inter-arrival times $\{X_i\}_{i\geq 1}$, and the indicator of the maintenance types are $\{\delta_i\}_{i\geq 1}$. $\delta_i = 0$ for CM and $\delta_i = 1$ for PM.

Five approximation models are considered, namely,

- HPP. The Homogeneous Poisson Process with a rate equaling the asymptotic rate of the SRP. This model is the simplest SRP approximation and is valid when the number of sources in the SRP tends to infinity [35].
- SIM. The Stationary Interval Method [105] corresponds to a renewal process whose intervals are distributed as in an SRP.
- IAT_1. The Inter-Arrival-Time copula of memory 1. This is a semi-parametric model that combines the generic distribution of the SRP with a parametric copula that models the correlation between two adjacent intervals. The IAT_k consists of a copula taking into account the joint distributions of the last k inter-arrival times.
- ARA_∞. The ARA_∞ process belongs to the family of virtual age models. It is a plausible candidate because two adjacent intervals in an ARA_∞ are negatively correlated, as in an SRP.
- BP. The Brown Proschan model assumes that maintenance is either perfect with probability p or imperfect with probability $1 - p$. Like an SRP, BP has a stationary regime. It is conceivable that SRP with heterogeneous components are relatively better approximated by BP, wherein the maintenance is heterogeneous.

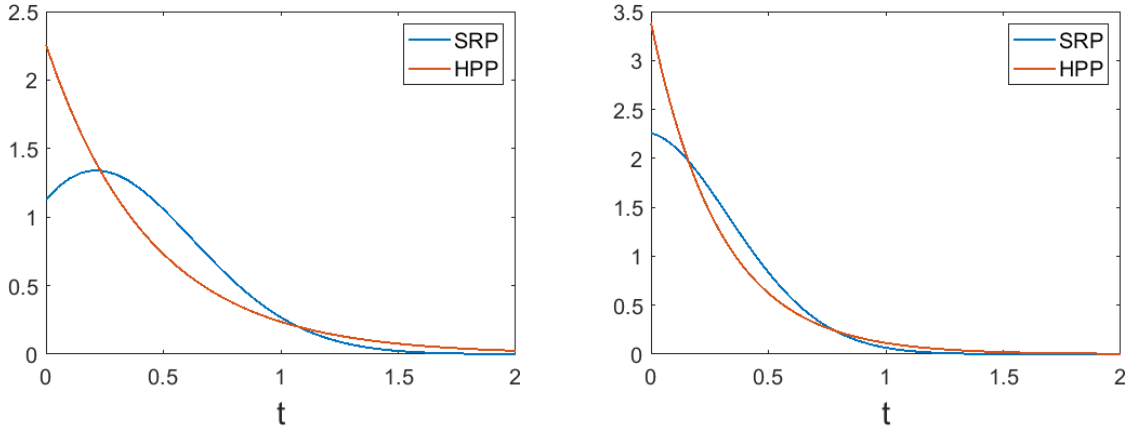
The use of IAT_1, ARA_∞ and BP to approximate an SRP is an original contribution. The reason why these approximations are considered will be elaborated later.

5.1.1 Homogeneous Poisson process

HPP has one parameter, i.e., the failure rate λ. Given the repair times $\{T_i\}_{i\geq 1}$ and repair types $\{\delta_i\}_{i\geq 1}$, λ can be assessed by MLE:

$$\widehat{\lambda} = \frac{n - \sum_{i=1}^{n} \delta_i}{T_n}. \tag{5.1}$$

The mean lifetime of the HPP, $\mathbb{E}[X_\infty]$, equals $1/\widehat{\lambda}$. The correlation between intervals, $Corr$, is 0 since the intervals are independent in an HPP. Although HPP will generally not cause errors in the mean lifetime, it is not perfect: take, for example, the Pdf of the lifetime in the steady-state. Consider the superposition of two Weibull renewal process with common scale 1 and shape 2, with a total of 10^6 successive failure times. The HPP parameter is estimated as $\hat{\lambda} = 0.4434$. It can be observed in Figure 5.1a that the distance between the Pdf of X_∞ in an SRP (drawn with blue curves) and that in its HPP approximation (red curve) is significant. When three renewal processes (with $\eta = 1$ and $\beta = 2$) are superimposed, $\hat{\lambda} = 0.2955$, and the distance between the densities is reduced, as is shown in Figure 5.1b.



(a) $\eta = [1, 1]$, $\beta = [2, 2]$ (b) $\eta = [1, 1, 1]$, $\beta = [2, 2, 2]$

Figure 5.1: Pdf of the cycle X_∞ in an SRP and in the approximating HPP.

5.1.2 Stationary interval method

The SIM method approximates an SRP by a renewal process whose interval is distributed identically as X_∞. Although the exact distribution of X_∞ is given by Eq.(2.72), we do not attempt to estimate its parameters because there will be too many parameters if the superimposed RP are different, and the statistical power will be weak if the dataset is small. Instead, this distribution is estimated by the Kaplan-Meier estimator [63], a widely applied non-parametric estimation approach:

$$\widehat{S(t)} = \prod_{i:t_i \leq t} 1 - \frac{d_i}{n_i} \tag{5.2}$$

with t_i a time when at least one event happened, d_i the number of events (e.g., failures) that happened at time t_i, and n_i the individuals known to have survived (have not yet had an event or been censored) up to time t_i.

To evaluate the accuracy of SIM approximation, $\mathbb{E}[X_\infty]$ is computed by numerically integrating a piecewise-linear version of the estimated survival function, and the correlation is 0 given that successive inter-arrival times are independent. Within the course of the inference procedures, additional non-parametric estimations of the reliability function have been tested using different kernels and several bandwidths, but the piecewise-linear version of the Kaplan-Meier has been chosen as it consistently provided the best results.

In the presence of PM, the Kaplan-Meier estimator is not defined beyond the largest observation if the longest interval is censored, making it impossible to compute the expected value. Therefore, we adopt the conventional approach, which consists of changing the largest observation to a death time if it is censored [36].

5.1.3 IAT_1 model

The IAT_1 is a refined version of SIM: it takes into account additionally the dependence between two successive intervals in the stationary regime by defining a copula. Copulas are an important part of the study of dependence between two variables since they allow us to separate the effect of dependence from the effects of the marginal distributions. In our case, for an SRP, the study of the dependence structure between two successive intervals can, therefore, be done independently from the marginal distributions by adopting a semi-parametric approach wherein the marginal survival function is assessed in the same way as the SIM method using a Kaplan-Meier estimator, and the dependence is modeled by a Frank copula. The latter is characterized by a single parameter θ. Both negative dependency ($\theta < 0$), positive dependency ($\theta > 0$) and independence ($\theta = 0$) can be modeled within the Frank family. Its distribution is given below:

$$C_\theta(x, y) = -\frac{1}{\theta} log\left(1 + \frac{(e^{-\theta x} - 1)(e^{-\theta y} - 1)}{e^{-\theta} - 1}\right). \tag{5.3}$$

Before explaining why we use particularly the Frank copula instead of Gaussian, t, or others, we introduce first the notion of the copula.

5.1.3.1 Copula and Sklar's theorem

In probability theory and statistics, a copula is a multivariate cumulative distribution function for which the marginal probability distribution of each variable is uniform on the interval [0, 1]. Consider a random vector $(X_1, X_2, \ldots, X_d)$ with continuous marginals distributions $F_i(x) = P(X_i \leq x)$. By applying the probability integral transform to each component, the random vector

$$(U_1, U_2, \ldots, U_d) = (F_1(X_1), F_2(X_2), \ldots, F_d(X_d)),$$

has marginals that are uniformly distributed on the interval [0, 1]. The copula of $(X_1, X_2, \ldots, X_d)$ is defined as the joint cumulative distribution function of $(U_1, U_2, \ldots, U_d)$:

$$C(u_1, u_2, \ldots, u_d) = P(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_d \leq u_d). \tag{5.4}$$

The copula C contains all information on the dependence structure between the components of $(X_1, X_2, \ldots, X_d)$ whereas the marginal cumulative distribution functions F_i contain all information on the marginal distributions.

The role that copulas play in the relationship between bivariate distribution functions and their univariate marginals is explained by Sklar's theorem [99]:

Theorem 5.1 *Let H be a joint distribution with marginals F and G. Then, there exists a copula C such that, for all x, $y \in [-\infty, \infty]$,*

$$H(x, y) = C(F(x), G(y)). \tag{5.5}$$

In other words, a bivariate copula is simply the uniform representation of the joint distribution in question. Recall that for an SRP, the bivariate distribution of two adjacent intervals is given by Eq.(2.73). Deriving the exact copula function from Eq.(2.73) is not apparent. Instead, we attempt to estimate the marginal distribution and the copula from the SRP pooled output empirically.

5.1.3.2 Justification of the selection of the Frank copula.

The choice of the Frank family is explained as follows. For illustration, consider the following four configurations of the SRP:

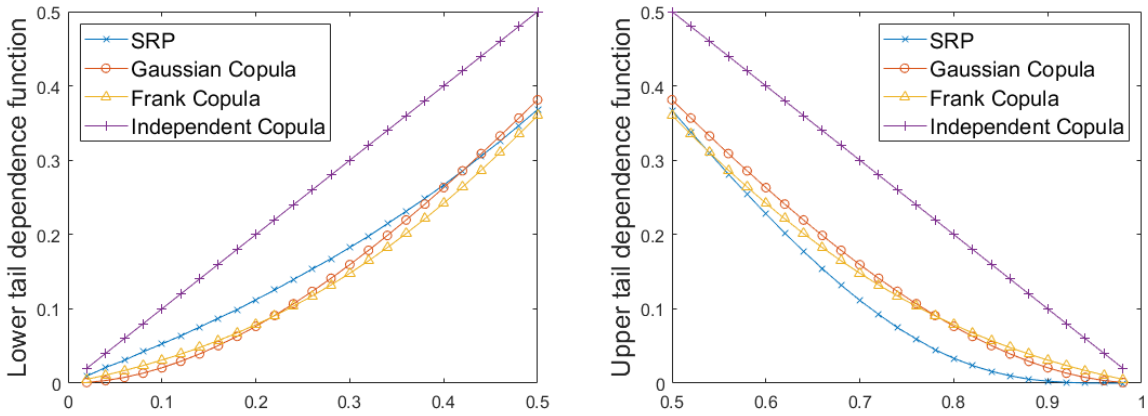
- I: a system formed of 3 homogeneous components with a low wear-out rate: $\boldsymbol{\eta} = [1, 1, 1], \boldsymbol{\beta} = [1.5, 1.5, 1.5]$.
- II: a system formed of 3 homogeneous components with a fast wear-out rate: $\boldsymbol{\eta} = [1, 1, 1], \boldsymbol{\beta} = [3.5, 3.5, 3.5]$.
- III: a system formed of 3 heterogeneous components with a low wear-out rate: $\boldsymbol{\eta} = [1, 2, 10], \boldsymbol{\beta} = [1.5, 1.5, 1.5]$.
- IV: a system formed of 3 heterogeneous components with a fast wear-out rate: $\boldsymbol{\eta} = [1, 2, 10], \boldsymbol{\beta} = [3.5, 3.5, 3.5]$.

For each configuration, an SRP sequence of length $N = 500000$ is generated, and several copula families (Gaussian [48], t [27], Frank, Gumbel, Clayton and independent copula) are then fitted to the data. It has been found that there exists a certain convergence between the families:

- The degree of freedom ν in the t-copula is systematically estimated to be extremely large, corresponding to the Gaussian copula. Therefore only the Gaussian copula is maintained in the comparative study.

- The parameter of the Gumbel family, $\widehat{\theta}_{\text{Gumbel}}$ is evaluated to be 1, and that of the Clayton family, $\widehat{\theta}_{\text{Clayton}}$ is estimated to be 0_+, which corresponds in both cases to the independent copula. As the Gumbel and Clayton copulas do not account for the negative dependency, both models are not kept for the comparative study.

Since the correlation between the intervals is of interest to the current paper, the independent copula, as well as the Gumbel and Clayton, are excluded from consideration. Gaussian and Frank copula are now the two candidates. In the following, we investigate whether they differ in estimating the tail dependence.



(a) Lower tail dependence function (b) Upper tail dependence function

Figure 5.2: An example of tail dependence functions estimated by different copulas.

The tail dependence concept describes the amount of dependence in the lower-left-quadrant tail or upper-right-quadrant tail of a bivariate distribution. Let F_1 and F_2 be the marginal distributions of variables X_1 and X_2. A common measure of tail dependence is given by the so-called upper/lower tail-dependence function [47]:

$$\lambda_L(v) = P(F_1(X_1) \leq v | F_2(X_2) \leq v) = \frac{C(v,v)}{v}, \tag{5.6}$$

$$\lambda_U(v) = P(F_1(X_1) \geq v | F_2(X_2) \geq v) = \frac{1 - 2v + C(v,v)}{1 - v}. \tag{5.7}$$

An example of the tail dependence functions $\lambda_L(v), v \in [0, 0.5]$ and $\lambda_U(v), v \in [0.5, 1]$ are shown respectively in Figures 5.2a and 5.2b for the SRP configuration II. The tail dependence of the intervals in SRP is estimated empirically [16]. It is observed that the Frank (triangle marker) and the Gaussian copulas (circle marker) are both close to the empirical SRP (cross marker). The independent copula is clearly inadequate.

SRP	θ in Frank Copula	ρ in Gaussian Copula	GOF Frank	GOF Gaussian
I	-0.7372	-0.1234	0.949	0.966
II	-2.3652	-0.3634	0.926	0.952
III	-0.6137	-0.1046	0.955	0.963
IV	-2.0441	-0.3151	0.926	0.946

Table 5.1: Estimated Copula parameters and results of test of goodness of fit.

The goodness-of-fit tests [49] are then used to assess to what extent the copula in an SRP resembles a Frank copula or a Gaussian copula. The copula under the null hypothesis is Frank or Gaussian. Let α be the level of significance. With unlimited data, an SRP copula can never be approximated by a Frank/Gaussian copula, meaning that the null hypothesis will be 100 percent rejected (p value less or equal to α). When the number of observations is limited, however, the SRP copula can be estimated very satisfyingly by a Frank/Gaussian copula.

Let $\alpha = 0.05$. An SRP sequence of length $N = 20$ is generated 1000 times, and the corresponding copula (which describes the dependence structure between two successive intervals in an SRP) is tested against Frank/Gaussian, using the Cramer-Von Mises statistics [49]. The total proportion of an SRP copula being classified as Frank/Gaussian is recorded in Table 5.1, column "GOF Frank" and "GOF Gaussian." It is found that the proportion of SRP copula, which "look like" a Gaussian copula, is slightly higher than that for a Frank copula.

To sum up, both Frank and Gaussian copula have similar performance in estimating the tail dependence. In terms of goodness of fit test, Gaussian is slightly better than Frank. Nevertheless, Frank copula admits an explicit formula, which is not possible for Gaussian. This facilitates the inference procedure greatly when missing/censored data is involved, or when PM is implemented, as shown later in this chapter.

5.1.3.3 Parameter estimation

Suppose that we have observed $n+1$ successive intervals, $X_1, X_2...X_{n+1}$ in an SRP when it has entered its steady state. To determine Frank copula that describes the dependence structure, we first decompose the observation sequence into two vectors:

$$\boldsymbol{U} = \{U_1, U_2...U_n\} = \{X_1, X_2...X_n\}, \tag{5.8}$$

and

$$\boldsymbol{V} = \{V_1, V_2...V_n\} = \{X_2, X_3...X_{n+1}\}. \tag{5.9}$$

Note that $\boldsymbol{U}$ and $\boldsymbol{V}$ are the interval lengths and take value from $\mathbb{R}^+$. Since we are in the stationary state, the marginal distributions for $\boldsymbol{U}$ and $\boldsymbol{V}$ are the same: both could be described by Eq.(2.72) and are empirically estimated by the Kaplan-Meier estimator. In order to estimate parametrically the Frank copula, we need to transform $\boldsymbol{U}$ and $\boldsymbol{V}$ into "pseudo-observation", $F_u(U_i)$ and $F_v(V_i)$, that are defined on $[0, 1]$. The functions F_u and

F_v are defined as:

$$F_u(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbb{1}(U_i \leq x), \quad F_v(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbb{1}(V_i \leq x), \tag{5.10}$$

where the factor $n+1$ (instead of n) allows the avoidance of boundary problems. The quantities $F_u(U_i)$ and $F_v(V_i)$ are the ranks of U_i and V_i, strictly less than 1.

Given the Frank copula assumption and the pseudo-observations $F_u(U_i)$ and $F_v(V_i)$, θ can be estimated from classic maximum likelihood estimation. The density of the Frank copula is given by

$$c_\theta(x,y) = \frac{\partial^2}{\partial x \partial y} C_\theta(x,y) = \frac{\theta e^{-\theta(x+y)}(1-e^{-\theta})}{[e^{-\theta} - 1 + (e^{-\theta x}-1)(e^{-\theta y}-1)]^2}, \tag{5.11}$$

with the likelihood

$$L(\theta | U_i, V_i) = \prod_{i=1}^{n} c_\theta(F_u(U_i), F_v(V_i)). \tag{5.12}$$

When preventive maintenance is implemented, the pairs (U_i, V_i) may be censored: the system is still working at the repair instant, and it is no longer appropriate to multiply the density functions. Let $\delta_i, i \in 1...n+1$ be the maintenance types with $\delta_i = 0$ for CM and $\delta_i = 1$ for PM. Four scenarios are to be considered:

(1). $\delta_i = 0$ and $\delta_{i+1} = 0$. The two adjacent intervals U_i and V_i are failure times, we can apply directly the density function:

$$M(\theta, x, y | \delta_i = 0, \delta_{i+1} = 0) = c_\theta(x,y). \tag{5.13}$$

(2). $\delta_i = 1$ and $\delta_{i+1} = 0$. The first interval is censored. We shall replace the density by the partial derivative

$$M(\theta, x, y | \delta_i = 1, \delta_{i+1} = 0) = \frac{\partial}{\partial y} C_\theta(x,y) = \frac{e^{-\theta y}(e^{-\theta x}-1)}{e^{-\theta} - 1 + (e^{-\theta x}-1)(e^{-\theta y}-1)}. \tag{5.14}$$

(3). $\delta_i = 0$ and $\delta_{i+1} = 1$. The first interval is a failure time and the second interval is censored. We shall replace the density by

$$M(\theta, x, y | \delta_i = 0, \delta_{i+1} = 1) = \frac{\partial}{\partial x} C_\theta(x,y) = \frac{e^{-\theta x}(e^{-\theta y}-1)}{e^{-\theta} - 1 + (e^{-\theta x}-1)(e^{-\theta y}-1)}. \tag{5.15}$$

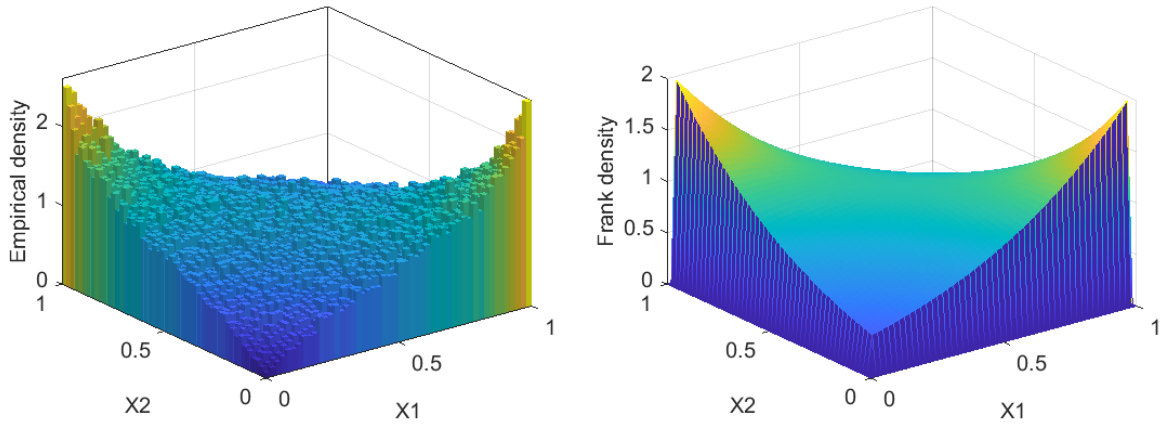
(4). $\delta_i = 1$ and $\delta_{i+1} = 1$. The two intervals are censored. We shall replace the density by the distribution function:

$$M(\theta, x, y | \delta_i = 1, \delta_{i+1} = 1) = C_\theta(x,y). \tag{5.16}$$

Finally, the likelihood with censored data is given by:

$$L(\theta | U_i, V_i, \delta_i) = \prod_{i=1}^{n} M(\theta, F_u(U_i), F_v(V_i) | \delta_i, \delta_{i+1}). \tag{5.17}$$

Compare now the empirical copula in an SRP and its Frank fit. Two identical Weibull renewal processes form the SRP with scale 1 and shape 2. With a total of 10^6 successive failure times, the Frank copula parameter is estimated to be $\theta = -1.7736$. The empirical density and fitted Frank copula density are respectively plotted in Figures 5.3a and 5.3b.



(a) Empirical copula (b) Frank copula fit

Figure 5.3: The dependence structure in SRP and its Frank copula fit.

5.1.3.4 Correspondence between SRP and Frank copula

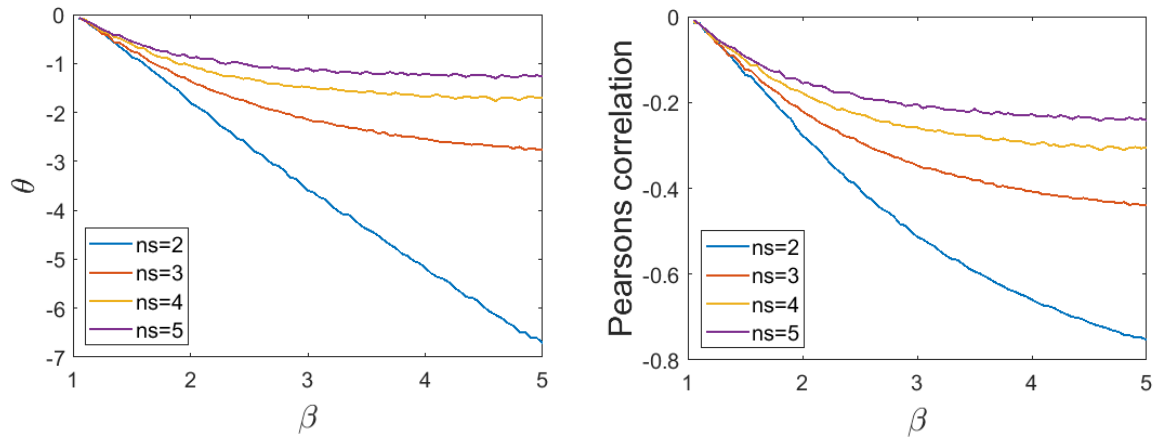
When the amount of observation is considerable, i.e., the SRP sequence has an infinite length, the SRP parameters uniquely define the parameter of the Frank copula fitted to the data. In other words, when fitting a Frank copula to the pooled output of the superposition of ns independent Weibull renewal processes with underlying parameters $\boldsymbol{\eta} = \eta_1, \eta_2 ... \eta_{ns}$ and $\boldsymbol{\beta} = \beta_1, \beta_2 ... \beta_{ns}$, there exists some function g: $\mathbb{R}_+^{2\times ns} \to \mathbb{R}$ such that

$$\theta = g(\boldsymbol{\eta}, \boldsymbol{\beta}). \tag{5.18}$$

We have not found the explicit form of the function g, and θ still needs to be estimated by maximizing the likelihood function defined in Eq.(5.17). Yet, the correspondence could easily be visualized by inferencing on a large SRP dataset. For illustration, consider the superposition of ns identical renewal processes, with common scale parameter η_c and shape β_c. η_c is set to 1. The number of renewal process $ns \in \{2, 3, 4, 5\}$, as we are interested in the case where relatively few renewal processes are superposed. $\beta_c \in 1 : 0.05 : 5$ since β is rarely more comprehensive than 5 in practice. $\hat{\theta}$ is plotted in Figure 5.4a, whereas the Pearson's correlation between two successive intervals in an SRP is shown in Figure 5.4b, for a given β_c and ns. The positive correlation between $\hat{\theta}$ and Pearson's correlation is evident.

5.1.4 ARA_∞ model

ARA_∞ is an alternate possibility to approximate an SRP. The idea is that both SRP and ARA_∞ have a stationary state: the failures arrive at a constant rate asymptotically.



(a) $\hat{\theta}$ (b) Pearson's correlation

Figure 5.4: $\hat{\theta}$ of the Frank copula (left) and Pearson's correlation between two successive intervals in an SRP (right).

Particularly, they exhibit a negative correlation between two adjacent intervals when the baseline failure rate is an increasing function of time: the larger X_n is, the smaller X_{n+1} may be. On the one hand, for ARA_∞, the larger X_n is, the larger A_n will be, which means the system is "older". Thus the next interval X_{n+1} may be small. On the other hand, for an SRP, the larger X_n is, the more other components are aging (except for the one that is renewed at the n-th repair), so the next failure may arrive soon. This is also illustrated by the negative values of θ of the Frank copula fitted to the SRP observation, as shown in Table 5.1. The mathematical modeling and comparison of the negative dependence are given below.

5.1.4.1 Negative dependence

The negative dependence is described by the Reverse Regular of Order 2 (RR2), stronger than the famous negative quadrant dependent (NQD). A pair of real-valued random variables (X_1, X_2) and its density function $f(\cdot,\cdot)$ are called RR_2 dependent [78] if

$$f(x_1, y_1)f(x_2, y_2) \leq f(x_2, y_1)f(x_1, y_2), \tag{5.19}$$

whenever $x_1 > x_2$ and $y_1 > y_2$. According to [64], this is equivalent to :

$$f(x, y)\frac{\partial^2 f}{\partial x \partial y} \leq \frac{\partial f}{\partial x}\frac{\partial f}{\partial y}. \tag{5.20}$$

Proposition 4 *Two successive intervals in an SRP, formed by identical components with IFR, are RR_2 dependent.*

proof 8 *We introduce the following corollary when the superimposed RP are identical:*

Corollary 5.1.1 *Note R,f the survival function and density function of an interval in the superposed RP. define $\phi(t) = \int_t^\infty R(u)du$. Then two successive intervals in the SRP*

in stationary regime is RR_2 dependent if and only if

$$\phi(t)f(t) \leq R(t)^2. \tag{5.21}$$

The proof is straightforward. Inequality 5.21 is easily satisfied in the case IFR since

$$\lambda(t)\int_t^\infty e^{-\Lambda(u)}du \leq \int_t^\infty \lambda(u)e^{-\Lambda(u)}du = R(t),$$

which proves that SRP is RR_2 dependent.

Proposition 5 *Two successive inter-arrival times of an ARA_∞ process with increasing baseline failure intensity are RR_2 dependent.*

proof 9 *The joint survival of two successive intervals in ARA_∞ conditioned on the previous virtual age satisfies equation 5.20, because*

$$\begin{array}{l} R(x,y) = P(X_{n+1} \geq y, X_n \geq x | A_{n-1} = a) \\ = \int_x^\infty P(X_{n+1} \geq y | X_n = u, A_{n-1} = a) f_{X_n|A_{n-1}=a}(u)du \\ = \int_x^\infty P(X_{n+1} \geq y | A_n = (1-\rho)(a+u))(-\frac{d}{du}P(X_n \geq u | A_{n-1} = a))du. \end{array}$$

Let $R_a(t) = P(X_n \geq t | A_{n-1} = a)$ and $R_u^b(t) = P(X_{n+1} \geq t | A_n = (1-\rho)(a+u))$. The derivative of $R(x,y)$ with respect to x, y and mixed derivative are respectively:

$$\frac{\partial R}{\partial x} = -R_x^b(y)(-\frac{d}{dx}R_a(x)),$$

$$\frac{\partial R}{\partial y} = \int_x^\infty \frac{d}{dy}R_u^b(y)(-\frac{d}{du}R_a(u))du,$$

$$\frac{\partial^2 R}{\partial x \partial y} = -\frac{d}{dy}R_x^b(y)(-\frac{d}{dx}R_a(x)).$$

The derivative of survival function $R_x^b(t)$is the product of failure rate and survival: $-\frac{d}{dy}R_x^b(y) = \lambda_x^b(y)R_x^b(y)$. Therefore, R is RR2 if and only if

$$\int_x^\infty R_u^b(y)(-\frac{d}{du}R_a(u))du\lambda_x^b(y) \leq \int_x^\infty \lambda_u^b(y)R_u^b(y)(-\frac{d}{du}R_a(u))du. \tag{5.22}$$

Using mean value theorem, the r.h.s. is reformulated as

$$\lambda_s^b(y)\int_x^\infty R_u^b(y)(-\frac{d}{du}R_a(u))du, \quad s \in [x,\infty].$$

When the system is IFR, $\lambda_x^b(y) \leq \lambda_s^b(y)$, leading to Eq.(5.22). Therefore, ARA_∞ is RR_2 dependent.

Although SRP and ARA_∞ display the RR_2 negative dependence between adjacent intervals, we must point out that the dependence structure in an SRP is different from

that in an ARA_∞ process. Consider, for example, the situation where all the superimposed renewal processes are identical with IFR. Depending on the number of renewal processes, the correlation between *non-adjacent* SRP intervals may exhibit negative or positive dependence, as displayed later. For ARA_∞ with an increasing baseline failure intensity, however, this is not true: any two intervals in an ARA_∞ process are negatively dependent.

To investigate the dependence in ARA_∞ intervals, we introduce first the notions of *Positively Regression Dependent* and *Negatively Regression Dependent* [78].

Definition 5.1 *Y is said to be positively regression dependent in X, if $P(Y > y|X = x)$ is increasing in x, for all y. Conversely, Y is said to be negatively regression dependent in X, if $P(Y > y|X = x)$ is decreasing in x, for all y.*

The following proposition shows that the virtual ages after *any* two repairs in an ARA_∞ are PRD dependent.

Proposition 6 *Denote by A_n the virtual age after the n-th repair in an ARA_∞ process. $\forall n = 1, 2, ..., k = 1, 2, ...$, A_{n+k} is stochastically increasing (SI) in A_n: $P(A_{n+k} > t|A_n = a)$ is increasing in a.*

proof 10 *Let $R_a(t) = P(Z > t|Z > a)$. When $k = 1$,*

$$P(A_{n+1} > x|A_n = a) = 1 \cdot \mathbb{1}(x < (1-\rho)a) + R_a(\frac{x}{1-\rho} - a) \cdot \mathbb{1}(x > (1-\rho)a).$$

Obviously, $\forall a_1 < a_2$, $P(A_{n+1} > x|A_n = a_1) \leq P(A_{n+1} > x|A_n = a_2)$. $P(A_{n+1} > x|A_n = a)$ is thus increasing in a. Assume the corollary 6 holds for $K = k$:

$$\forall a_1 < a_2, P(A_{n+k} > x|A_n = a_1) \leq P(A_{n+k} > x|A_n = a_2).$$

We want to prove $P(A_{n+k+1} > x|A_n = a_1) \leq P(A_{n+k+1} > x|A_n = a_2)$. Let $R_1(t) = P(A_{n+k} > t|A_n = a_1)$ and $R_2(t) = P(A_{n+k} > t|A_n = a_2)$, $f_1(t) = -\frac{d}{dt}R_1(t)$ and $f_2(t) = -\frac{d}{dt}R_2(t)$, $F_1(t) = 1 - R_1(t)$ and $F_2(t) = 1 - R_2(t)$. Then,

$$\begin{aligned}
&P(A_{n+k+1} > s|A_n = a_1) = \int_0^\infty P(A_{n+k+1} > s|A_{n+k} = a) f_1(a) da \\
&= \Big[P(A_{n+k+1} > s|A_{n+k} = a)F_1(a)\Big]_0^\infty - \int_0^\infty [\frac{d}{da}P(A_{n+k+1} > s|A_{n+k} = a)]F_1(a)da \\
&= 1 - \int_0^\infty [\frac{d}{da}P(A_{n+k+1} > s|A_{n+k} = a)]F_1(a)da.
\end{aligned}$$

Similarly,

$$P(A_{n+k+1} > s|A_n = a_2) = 1 - \int_0^\infty [\frac{d}{da}P(A_{n+k+1} > s|A_{n+k} = a)]F_2(a)da.$$

$a_1 < a_2 \longrightarrow R_1(t) < R_2(t) \longrightarrow F_1(t) > F_2(t) \longrightarrow P(A_{n+k+1} > s|A_n = a_1) < P(A_{n+k+1} > s|A_n = a_2)$. This indicates that if Proposition 6 holds for $K = k$, then it holds for $K = k + 1$. Proof completed.

Having proved the PRD between any two virtual ages, we can now address the Negatively Regression Dependent (NRD) between any two intervals:

Proposition 7 *Denote by X_n the n-th interval in an ARA_∞ process. $\forall n = 1, 2, ..., k = 1, 2, ...$, X_{n+k} is negatively regression dependent in X_n: $P(X_{n+k} > t|X_n = x)$ is decreasing in x.*

proof 11 *Since $A_n = (1-\rho)(A_{n-1} + X_n)$ is increasing in X_n, A_{n+k} is thus positively regression dependent in X_n: $\forall x_1 < x_2, P(A_{n+k} > t|X_n = x_1) \leq P(A_{n+k} > t|X_n = x_2)$. Let $R_1(t) = P(A_{n+k} \geq t|X_n = x_1)$ and $R_2(t) = P(A_{n+k} \geq t|X_n = x_2)$, $f_1(t) = -\frac{d}{dt}R_1(t)$ and $f_2(t) = -\frac{d}{dt}R_2(t)$, $F_1(t) = 1 - R_1(t)$ and $F_2(t) = 1 - R_2(t)$. Then,*

$$
\begin{aligned}
&P(X_{n+k+1} > s|X_n = x_1) = \int_0^\infty P(X_{n+k+1} > s|A_{n+k} = a) f_1(a) da \\
&= \Big[P(X_{n+k+1} > s|A_{n+k} = a) F_1(a)\Big]_0^\infty - \int_0^\infty [\frac{d}{da} P(X_{n+k+1} > s|A_{n+k} = a)] F_1(a) da \\
&= -\int_0^\infty [\frac{d}{da} P(X_{n+k+1} \geq s|A_{n+k} = a)] F_1(a) da.
\end{aligned}
$$

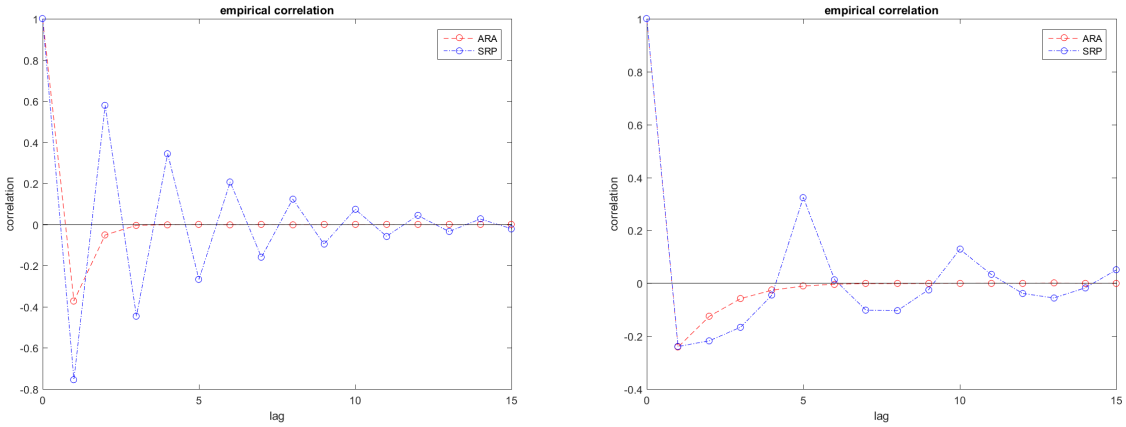
Similarly,

$$
P(X_{n+k+1} > s|X_n = x_2) = -\int_0^\infty [\frac{d}{da} P(X_{n+k+1} > s|A_{n+k} = a)] F_2(a) da.
$$

$x_1 < x_2 \longrightarrow R_1(t) < R_2(t) \longrightarrow F_1(t) > F_2(t) \longrightarrow P(X_{n+k+1} > s|X_n = x_1) > P(X_{n+k+1} > s|X_n = x_2)$. Therefore, $P(X_{n+k+1} > s|X_n = x)$ is decreasing in x.

To visualize the difference in the dependence structure of an SRP and of an ARA_∞, consider the SRP configured as $\eta_c = 1, \beta_c = 5$, $p = 2$. Generating an SRP observation of length $N = 10^6$, we can then derive the parameters of the approximating ARA_∞ by maximizing the likelihood defined in Eq.(5.24): $\hat{\alpha} = 0.254, \hat{\beta} = 5.505, \hat{\rho} = 0.375$. Generate now an ARA_∞ sequence of length N with $(\hat{\alpha}, \hat{\beta}, \hat{\rho})$. We can empirically obtain the auto-correlation of gap k: $c_k = corr(X_n, X_{n+k})$ is the Pearson's correlation between X_n and X_{n+k} in the SRP and in its ARA_∞ approximation.

In Figure 5.5a, the vertical axis represents the value of c_k while the horizontal axis stands for k. The correlation between intervals in the SRP is plotted with blue curve, while that in the approximating ARA_∞ is drawn with red curve. When $k = 0$, $corr(X_n, X_n) = 1$ for both SRP and ARA_∞; when $k = 1$, the correlations between two adjacent intervals are both negative, as stated in Propositions 4 and 5; when $k = 2$, the correlation between X_n and X_{n+2} in the ARA_∞ is still negative, whereas that in the SRP becomes positive. The same could be observed for the SRP configured as $\eta_c = 1, \beta_c = 5$, $p = 5$. This time, the parameters of the approximating ARA_∞ are $\hat{\alpha} = 0.572, \hat{\beta} = 5.406, \hat{\rho} = 0.155$, and the correlations between the intervals are plotted in Figure 5.5b: when when $k = 1$, the correlations between two adjacent intervals are both negative; but when $k = 5$, the correlation in SRP becomes positive while that in ARA_∞ is always negative. These observations are consistent with Proposition 7.



(a) $p = 2, \beta_c = 5$ (b) $p = 5, \beta_c = 5$

Figure 5.5: Pearson's correlation for SRP and its ARA_∞ approximation.

It is concluded here that although ARA_∞ is a plausible candidate to approximate an SRP, given that both of them have a steady-state and a negative correlation between adjacent intervals, ARA_∞ is not capable of imitating the dependence structure in an SRP fully, i.e., the correlation between X_n and X_{n+k}, which can sometimes be positive depending on its configuration.

5.1.4.2 Parameter estimation

Inference procedures have been implemented for virtual age models when the system, under CM, is assumed to be in its stationary regime at the beginning of the observations [28] for any given initial intensity. In the following, it is assumed that the initial intensity is given by the power law, which is a common assumption for aging systems in reliability engineering. Thus, three parameters, α, β (Weibull pseudo-scale and shape), and ρ_0 (the efficiency of CM), are to be estimated.

When both CM and PM are presented in the SRP, the ARA_∞ inference procedures need to be adapted since the maintenance is now heterogeneous. We shall integrate periodic PM into the ordinary ARA_∞ process: it is assumed that the PM efficiency follows an ARA_∞ assumption, i.e., reduces the virtual age by $(1 - \rho_1)$ where ρ_1 is likely in practice to be greater than ρ_0. The model parameters can be estimated by Maximum Likelihood Estimation. Denote by $\lambda(t) = \alpha\beta t^{\beta-1}$ the initial Weibull intensity and $\Lambda(t) = \alpha t^\beta$ the cumulative intensity. First, the likelihood function $\mathcal{L}_{a_0}$ associated with the observation of the n first maintenance times and types $(\mathbf{X}, \boldsymbol{\delta})_{\mathbf{1..n}}$ and an initial age a_0 can be expressed as in Eq.(5.23). The effective ages in Eq.(5.23) can be obtained by induction given a_0 and using $a_i = (1 - \rho_{\delta_i})(a_{i-1} + x_i)$.

$$\mathcal{L}_{a_0}((\mathbf{X}, \boldsymbol{\delta})_{\mathbf{1..n}}) = \prod_{i=1}^{n} \left(\lambda\left(a_{i-1} + x_i\right)\right)^{1-\delta_i} \times exp\left(\sum_{i=1}^{n} -\Lambda\left(a_{i-1} + x_i\right) + \Lambda\left(a_{i-1}\right)\right). \quad (5.23)$$

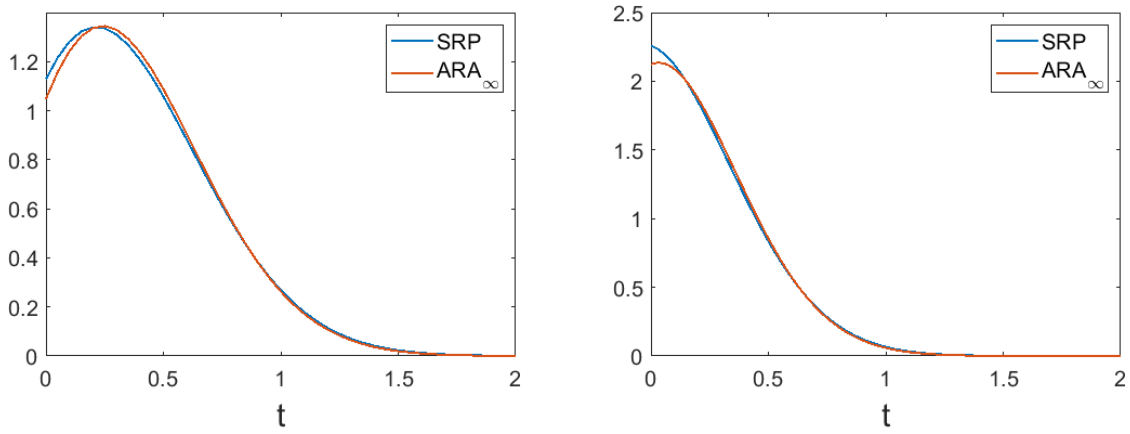
Second, as the SRP is assumed to be in its stationary regime at the beginning of the observations, the same assumption is proposed for its approximating model. It implies that

a_0 is the realization of the limiting age distribution A_∞^{ARA} with pdf $f_{A\infty}^{ARA}$. The resulting likelihood function $\mathcal{L}$ associated with the observation of the n first maintenance times and types $(\mathbf{X}, \boldsymbol{\delta})_{\mathbf{1..n}}$ and an initial mixing distribution A_∞^{ARA} is presented in Eq.(5.24). The mixing distribution A_∞^{ARA} has been characterized theoretically by Eq.(2.67) for CM only and can be derived empirically from intensive simulations considering CM and PM.

$$\mathcal{L}((\mathbf{X}, \boldsymbol{\delta})_{\mathbf{1..n}}) = \int_0^\infty \mathcal{L}_a((\mathbf{X}, \boldsymbol{\delta})_{\mathbf{1..n}}) f_{A_\infty}^{ARA}(a) da. \tag{5.24}$$

Finally, once the parameters of the model estimated, $\mathbb{E}[X_\infty]$ can be computed numerically by Eq.(2.70), and the correlation $Corr$ can be obtained via Monte Carlo simulation.

Consider the Pdf of the lifetime in the steady-state. When two Weibull renewal process with common scale 1 and shape 2 are superimposed, with a total of 10^6 successive failure times, the ARA_∞ parameters are estimated as $\hat{\alpha} = 1.85$, $\hat{\beta} = 2.075$, $\rho = 0.5992$. It can be observed in Figure 5.6a that the distance between the Pdf of X_∞ in an SRP (drawn with blue curves) and that in its HPP approximation (red curve) is insignificant. When three renewal processes (with $\eta = 1$ and $\beta = 2$) are superimposed, $\hat{\alpha} = 2.8386$, $\hat{\beta} = 2.0632$, $\rho = 0.4352$, and the distance between the densities is even smaller, as is shown in Figure 5.6b.



(a) $\eta = [1, 1]$, $\beta = [2, 2]$ (b) $\eta = [1, 1, 1]$, $\beta = [2, 2, 2]$

Figure 5.6: Pdf of the cycle X_∞ in an SRP and in the approximating ARA_∞.

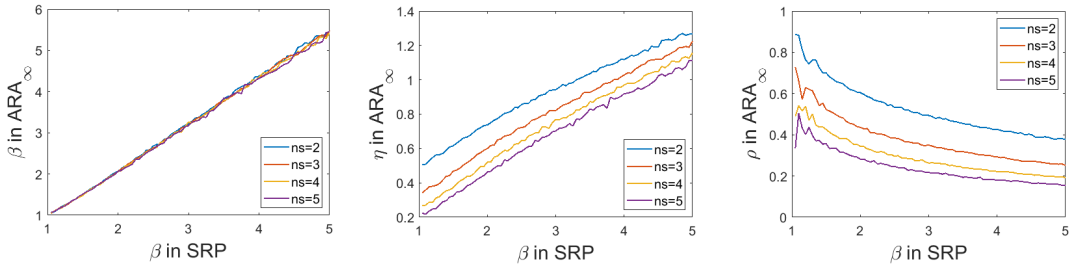
5.1.4.3 Correspondence between SRP and ARA_∞

Similarly to the Frank copula and SRP, there is a correspondence between the SRP sequence and its ARA_∞ approximation: when fitting an ARA_∞ process to the pooled output of the superposition of ns independent Weibull renewal processes with underlying parameters $\boldsymbol{\eta} = \eta_1, \eta_2 ... \eta_{ns}$ and $\boldsymbol{\beta} = \beta_1, \beta_2 ... \beta_{ns}$, there exists some function g: $\mathbb{R}_+^{2\times ns} \to \mathbb{R}_+^3$ such that

$$(\alpha, \beta, \rho) = g(\boldsymbol{\eta}, \boldsymbol{\beta}). \tag{5.25}$$

We have not found the explicit form of the function g, and the ARA_∞ parameters still need to be estimated by maximizing the likelihood function defined in Eq.(5.24). Yet,

the correspondence could easily be visualized by inferencing on a large SRP dataset. Consider the same SRP configurations: $\eta_c = 1$, $ns \in \{2, 3, 4, 5\}$, and $\beta_c \in 1 : 0.05 : 5$ The parameters of the approximating ARA_∞, $(\hat{\beta}, \hat{\eta}, \hat{\rho})$, are plotted in Figure 5.7.



(a) $\hat{\beta}$ (b) $\hat{\eta}$ (c) $\hat{\rho}$

Figure 5.7: Parameters of the approximating ARA_∞ as a function of β_c and of ns.

The linear relation between $\hat{\beta}$ and β_c is shown in Figure 5.7a. $\hat{\beta}$ is slightly larger than β_c, and the difference is more significant when β_c is large. The overlap of the curves shows that the number of renewal process has barely an influence on $\hat{\beta}$ when β_c is small.

Figure 5.7b shows that $\hat{\eta}$ increases with β_c and decreases with ns. In fact, when $\beta_c = 1$, $\hat{\eta} = 1/ns$: this is because the renewal processes become HPP. η_c in SRP and $\hat{\eta}$ in ARA_∞ are both proportional to the mean lifetime. Thus, when more renewal processes are superimposed, the mean lifetime of the SRP decreases, reducing the value of $\hat{\eta}$.

Figure 5.7c shows that $\hat{\rho}$ is a decreasing function of β_c and of ns. This is because the more components are superimposed, the less the proportion of a system-level renewal, which in turn results in a drop in maintenance efficiency. In addition, when $\beta = 1$, $\hat{\rho}$ is non-identifiable.

5.1.5 Brown-Proschan model

The Brown-Proschan model is another candidate to approximate an SRP as both converge to a stationary regime. Consider an initial Weibull intensity. Estimation methods have been presented in [79] and [76]. With only CM, the parameters to estimate are the scale and shape of the Weibull distribution and the probability of perfect CM, p_0. When both CM and PM are presented in the SRP, ordinary BP should be modified by incorporating periodic PM, which is perfect with probability p_1 and minimal with probability $1 - p_1$. The likelihood function associated with the observation of n maintenance times and types can be derived in three steps. First, the likelihood function $\mathcal{L}_{a_0}^{NHPP}$ associated with the observation of a system with initial age a_0 under minimal repair are reminded in Eq.(5.26) using the classic results of Non-Homogeneous Poisson processes [94]:

$$\mathcal{L}_{a_0}^{NHPP}((\mathbf{X}, \boldsymbol{\delta})_{\mathbf{1..n}}) = \prod_{i=1}^{n} \left(\lambda\left(a_0 + t_i\right)\right)^{1-\delta_i} \times exp\left(-\Lambda\left(a_0 + t_n\right) + \Lambda\left(a_0\right)\right). \tag{5.26}$$

Second, the likelihood function $\mathcal{L}_{a_0}^{BP}$ associated with the observation of a system with initial age a_0 and BP efficiencies are derived recursively in Eq.(5.27) using a similar

approach as in [28]:

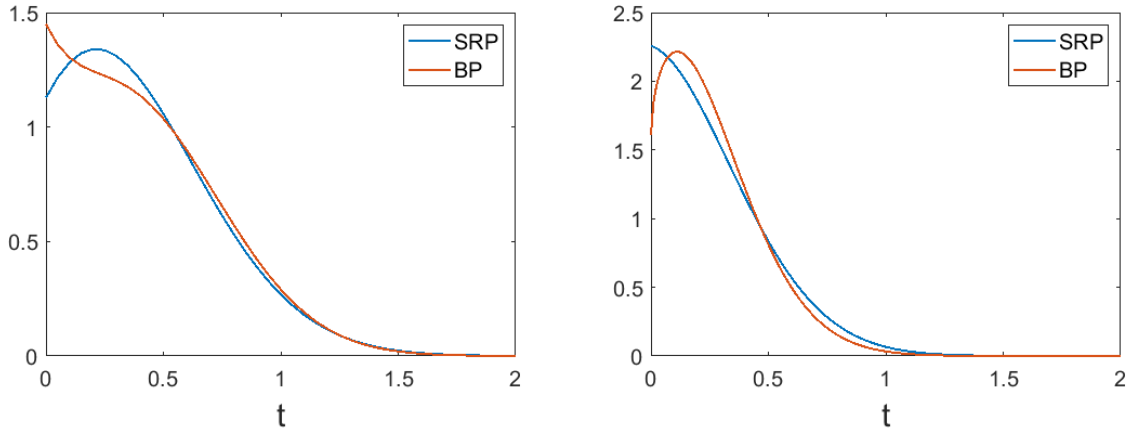
$$\mathcal{L}_{a_0}^{BP}((\mathbf{X},\boldsymbol{\delta})_{\mathbf{1..n}}) = \left[\prod_{k=1}^{n-1}(1-p_{\delta_k})\right] \times \mathcal{L}_{a_0}^{NHPP}((\mathbf{X},\boldsymbol{\delta})_{\mathbf{1..n}}) + \sum_{i=1}^{n-1} \mathcal{L}_{a_0}^{BP}((\mathbf{X},\boldsymbol{\delta})_{\mathbf{1..i}}) p_{\delta_i} \left[\prod_{k=i+1}^{n-1}(1-p_{\delta_k})\right] \times \mathcal{L}_{0}^{NHPP}((\mathbf{X},\boldsymbol{\delta})_{\mathbf{i+1..n}}). \quad (5.27)$$

Finally, the initial age a_0 is unknown and can be regarded as a realization of the random variable A_∞^{BP}, which is the asymptotic effective age in the stationary regime. The resulting likelihood function $\mathcal{L}^{BP}$ associated with the observation of the n first maintenance times and types $(\mathbf{X},\boldsymbol{\delta})_{\mathbf{1..n}}$ and A_∞^{BP} is given by

$$\mathcal{L}^{BP}((\mathbf{X},\boldsymbol{\delta})_{\mathbf{1..n}}) = \int_0^\infty \mathcal{L}_a^{BP}((\mathbf{X},\boldsymbol{\delta})_{\mathbf{1..n}}) f_{A_\infty^{BP}}(a)da. \quad (5.28)$$

The distribution of A_∞^{BP} is given by Eq.(2.39) with only CM. With CM and PM, the pdf of A_∞^{BP} can be obtained via simulation. Once the model parameters estimated, $\mathbb{E}[X_\infty]$ and $Corr$ can be obtained from a plugged-in version of their theoretical expressions or numerically.

Consider the Pdf of the lifetime in the steady-state. When two Weibull renewal process with common scale 1 and shape 2 are superimposed, with a total of 200 consecutive failure times, the BP parameters are estimated as $\hat{\alpha} = 2.5528$, $\hat{\beta} = 1.8938$, $\rho = 0.6381$. It can be observed in Figure 5.6a that the distance between the Pdf of X_∞ in an SRP (drawn with blue curves) and that in its HPP approximation (red curve) is significant for small values of t. When three renewal process (with $\eta = 1$ and $\beta = 2$) are superimposed, $\hat{\alpha} = 5.2058$, $\hat{\beta} = 1.6458$, $\rho = 0.6571$, and the densities are plotted in Figure 5.8.



(a) $\eta = [1,1]$, $\beta = [2,2]$ (b) $\eta = [1,1,1]$, $\beta = [2,2,2]$

Figure 5.8: Pdf of the cycle X_∞ in an SRP and in the approximating BP.

5.2 Performance evaluation

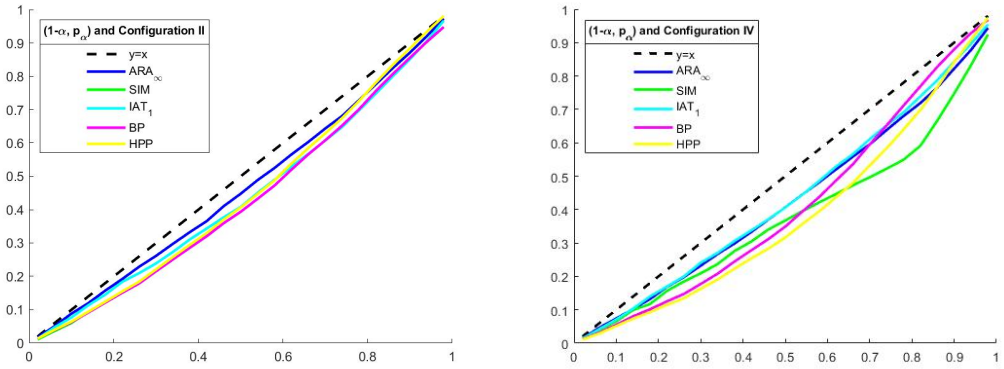
In this section, we propose two approaches to evaluate the performance of the approximating models. The first one is based on level set and is particularly useful when the parameters are known. The second one consists of accessing the approximation error in the mean lifetime and correlation between adjacent intervals. When the model parameters are unknown and have to be estimated, the amplitude of approximation error is dependent on the length of available data. For illustration, five SRP configurations are studied, and the lifetimes are supposed to be Weibull distributed:

- I: a system formed of 3 homogeneous components with a low wear-out rate: $\boldsymbol{\eta} = [1, 1, 1], \boldsymbol{\beta} = [1.5, 1.5, 1.5]$.
- II: a system formed of 3 homogeneous components with a fast wear-out rate: $\boldsymbol{\eta} = [1, 1, 1], \boldsymbol{\beta} = [3.5, 3.5, 3.5]$.
- III: a system formed of 3 heterogeneous components with a low wear-out rate: $\boldsymbol{\eta} = [1, 2, 10], \boldsymbol{\beta} = [1.5, 1.5, 1.5]$.
- IV: a system formed of 3 heterogeneous components with a fast wear-out rate: $\boldsymbol{\eta} = [1, 2, 10], \boldsymbol{\beta} = [3.5, 3.5, 3.5]$.
- V: a system formed of 6 relatively homogeneous components with a moderate wear-out rate: $\boldsymbol{\eta} = [1, 1, 1, 2, 2, 2], \boldsymbol{\beta} = [2.5, 2.5, 2.5, 2.5, 2.5, 2.5]$.

5.2.1 A level-set procedure

Let us consider a risk α and a trajectory of SRP observations where the parameters of the model and the age of each component are available. At time T_i, it is straightforward to compute the confidence interval of smallest amplitude $I = [T_i + a, T_i + b]$ with confidence level $100(1 - \alpha)\%$ from the actual model. Next, an interval $J = [T_i + a', T_i + a' + b - a]$ of amplitude $b - a$ can be determined which maximizes the confidence set of the approximating model. The probability that T_{i+1} belongs to I is naturally $1 - \alpha$ and the probability that T_{i+1} belongs to J is automatically less than $1 - \alpha$. The quality of the approximating model lies in how the latter probability is close to $1-\alpha$. This probability p_α can be estimated empirically from a unique and sufficiently long trajectory by assessing the proportion of times when the prediction has been correct. Furthermore, the approximated model's overall performance can be synthesized by its Gini index $G = 2\int_0^1 (1-\alpha-p_\alpha)d\alpha$. If $G = 0$, the approximating model is indistinguishable from the actual model, and $G = 1$ is the worst-case scenario where the approximating model is totally inconsistent with SRP.

For each model (HPP, SIM...BP) and each configuration (I, II...V), the level set p_α is computed and compared to the initial confidence level $100(1 - \alpha)\%$ through the identity line. Figure 5.9a and 5.9b present the results with a fast wear-out for the configurations II and IV, respectively. The figures for other configurations are not presented as the plots are almost indistinguishable from the identity line. In addition, each Gini coefficient is provided in Table 5.2.



(a) Homogeneous case (Conf. II) (b) Heterogeneous case (Conf. IV)

Figure 5.9: Level sets in the fast aging rate cases.

	HPP	SIM	IAT_1	ARA_∞	BP
I	9.1	8.9	8.5	8.2	8.6
II	119.2	119.5	137.0	89.4	144.0
III	6.5	7.2	6.2	3.0	5.9
IV	222.7	230.1	116.1	133.8	204.5
V	2.8	4.3	4.3	3.0	5.6

Table 5.2: Gini coefficients (10^{-3}).

The analysis of the level sets and Gini is given below:

- For each model, the Gini index is relatively low, between 0.002 and 0.2, which indicates that each model provides an efficient approximation of an SRP. The Gini index is particularly low when the components' aging is moderate or when the number of components is important.
- The model under ARA_∞ assumption almost consistently provides the best index, regardless of the heterogeneity of the components and the aging rates of the components.
- The IAT_1 and BP models present satisfactory but more contrasting results. The BP approximation is comparatively much less efficient when the components are aging fast (Configuration II and IV). The IAT_1 approximation outperforms the other models in the heterogeneous case with fast aging rates (Configuration IV) but provides otherwise similar results to the HPP approximations.
- The two static models HPP and SIM present decent and very similar results. When the aging rate is moderate, their efficiency has the same order of magnitude as the other adaptive models, but they are outperformed by ARA_∞ when the wear-out rate is fast.
- All the models are extremely efficient when the number of components is significant, particularly the HPP, which is the theoretical limiting case when the number of components tends to infinity [35].

5.2.2 Mean and correlation

The parameters of the models are commonly not known and need to be estimated first from a relatively small dataset. The number of observations is assumed to be larger than 10 without exceeding 200, and the system is assumed to be in its stationary regime at the beginning of the observation. Three configurations previously studied in Section 5.2.1 are considered, two with only CM and one with both CM and periodic PM:

- A. The configuration II.
- B. The configuration IV.
- C. The configuration II with periodic PM. The periodicity of the preventive maintenance policy is $\Delta = 0.425$. The periodicity has been chosen so that the proportion of PM in the actual SRP model is 70%. The maintenance efficiency is assumed to be imperfect but of great quality on each component. Therefore, it has been opted to assume that PM efficiencies are ARA_∞ with an improvement factor $\rho = 0.7$. The virtual age of each component after a preventive maintenance action is 30% of its virtual age just before the PM.

These three configurations are not exhaustive but are quite representative of the behavior of an SRP with or without homogeneity. The PM policy in (C), which corresponds to the case where imperfect PM is carried out simultaneously on all the components, is investigated because it is easy to implement and widely applied in practice.

Two predictors, the mean lifetime $\mathbb{E}[X_\infty]$ and the Pearson's correlation coefficient $Corr$ of two successive inter-failure times of an SRP in its stationary state, have been considered.

These two predictors are assumed to be known for the SRP models. They can be obtained theoretically considering only CM or based on Monte Carlo simulations at any given precision when both CM and periodic PM are implemented. Given a sample size N of an SRP configuration, the mean-squared error (MSE) of the two predictors $\mathbb{E}[X_\infty]$ and $Corr$ are derived empirically for the five approximated models based on 5000 replicates of histories. A general structure of the simulation procedure is presented in the Algorithm 2.

The results are as follows: the MSE of the two predictors $\mathbb{E}[X_\infty]$ and $Corr$ are respectively plotted in Figure 5.10a,5.10b for configuration A, 5.11a,5.11b for configuration B and 5.12a,5.12b for configuration C. The MSE of the two predictors $\mathbb{E}[X_\infty]$ and $Corr$ are respectively plotted in Figures 5.10a,5.10b for configuration A, 5.11a,5.11b for configuration B and 5.10a,5.10b for configuration C.

When no PM is involved (Conf. A and B), HPP has the smallest error in the estimation of $\mathbb{E}[X_\infty]$ (see Figures 5.10a and 5.11a). Virtual age models (ARA_∞ and BP) perform better than IAT_1 and SIM. With more than 50 data, the differences between the MSEs given by the tested models are less important.

As for the error of the predictor $Corr$ (see Figures 5.10b and 5.11b), HPP and SIM have the largest error since their intervals are independent. BP is an auto-correlated

Algorithm 2 Empirical MSE computation

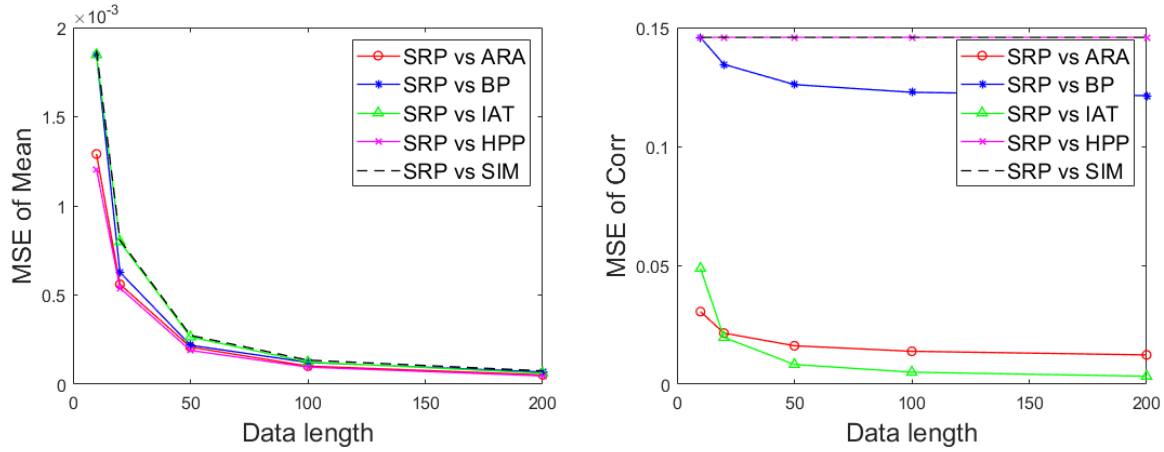
Select a SRP configuration
Compute the predictors $e = \mathbb{E}[X_\infty]$ and $r = Corr$
for $N \in \{10, 20, 50, 100, 200\}$ **do** *//N=length of a trajectory*
  for $k = 1 : 5000$ **do** *//kth sample*
    Simulate a SRP trajectory **Hist** of length N
    for $i = 1 : 5$ **do** *//ith model (HPP, SIM, IAT$_1$, ARA$_\infty$, BP)*
      Assess the parameters of the ith model from **Hist**
      Estimate the predictors by $\widehat{e_i(k)}$ and $\widehat{r_i(k)}$
    end for
  end for
  for $i = 1 : 5$ **do**
    Compute the empirical MSE of each measure for a trajectory of length N and the approximated model i:
    $\widehat{MSE_E(i, N)} = 1/5000 \sum_{k=1}^{5000} (\widehat{e_i(k)} - e)^2$
    $\widehat{MSE_{Corr}(i, N)} = 1/5000 \sum_{k=1}^{5000} (\widehat{r_i(k)} - r)^2$
  end for
end for

process with a weak dependence between intervals, but its estimation of correlation is far from accurate. IAT_1 has the smallest MSE when the superimposed RP is homogeneous, which may support the practice of using a Frank copula to approximate the dependence structure in such an SRP. If the superimposed RP differ strongly from each other as in configuration B, ARA_∞ outperforms other models in estimating the correlation between adjacent intervals.

When periodic PM are implemented (Figures 5.12a and 5.12b), the SIM and IAT_1 have a considerable error in $\mathbb{E}[X_\infty]$. HPP performs the best when estimating $\mathbb{E}[X_\infty]$, but with more data, the advantage of HPP over BP/ARA_∞ is less significant. As for the correlation estimator, the performance of the tested models depends on the data length: with less than 50 data, HPP outperforms the others; otherwise, BP and ARA_∞ are the best. It should be emphasized that the implementation of periodic PM results in a non-null correlation in HPP/SIM sequences.

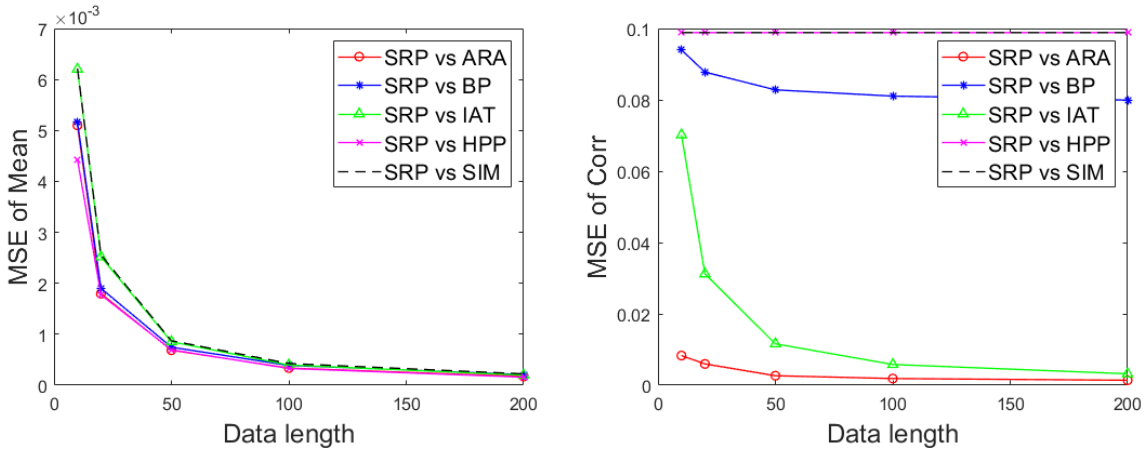
5.3 Conclusion

In this chapter, we have introduced and compared five approximating models for an SRP: two imperfect repair models (ARA_∞ and Brown-Proschan), two renewal processes (SIM and HPP) and IAT_1, constructed by the Kaplan-Meier estimated marginal distribution and a Frank copula which captures the dependence structure between successive intervals. The performances of these models are evaluated by investigating the amplitude of errors of mean interval length and correlations when the approximations mentioned above are used. Further, their capabilities in the prognosis of RUL are examined using the level-set approach and Gini index. It is difficult to say which model is overall the best: their performances depend on the aging rate as well as the available data amount.



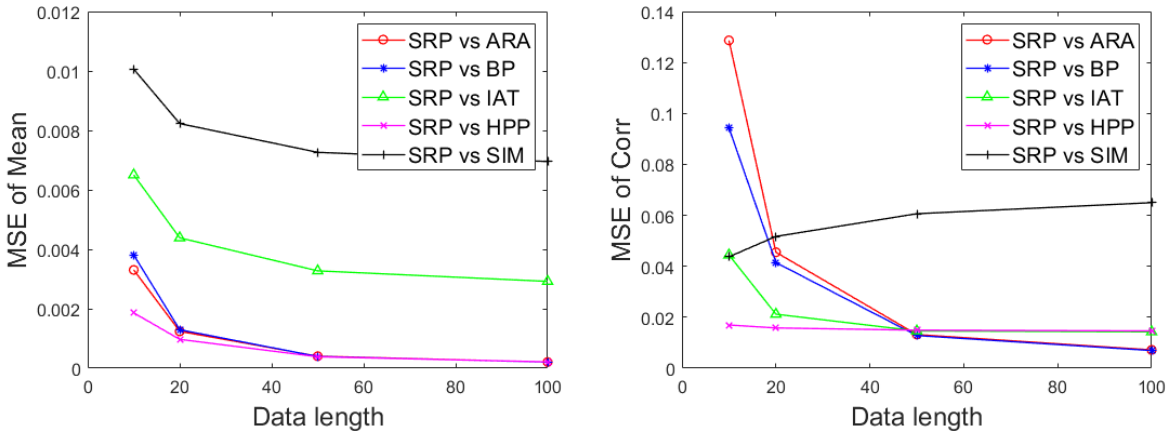
(a) MSE of $\mathbb{E}[X_\infty]$ (b) MSE of *Corr*

Figure 5.10: Configuration A: $\boldsymbol{\eta} = [1, 1, 1], \boldsymbol{\beta} = [3.5, 3.5, 3.5]$.



(a) MSE of $\mathbb{E}[X_\infty]$ (b) MSE of *Corr*

Figure 5.11: Configuration B: $\boldsymbol{\eta} = [1, 2, 10], \boldsymbol{\beta} = [3.5, 3.5, 3.5]$.



(a) MSE of $\mathbb{E}[X_\infty]$ (b) MSE of *Corr*

Figure 5.12: Configuration C: $\boldsymbol{\eta} = [1, 1, 1], \boldsymbol{\beta} = [3.5, 3.5, 3.5]$ with periodic PM.

The benefits of using these approximation approaches will be demonstrated in the next chapter, where failure data collected from the Norwegian railway signaling system shall be investigated. We show how these models could be used to evaluate the system's reliability and the remaining useful life, and highlight particularly the gain of using imperfect repair models that conserve the negative dependence between intervals in an SRP.

Chapter 6

Bane NOR failure data analysis

Previously in Chapters 3 and 4, we have investigated the asymptotic properties and the heterogeneity of imperfect repair models. Specifically, in Chapter 5, we have addressed how these models could approximate the superposition of renewal processes–a common modeling approach for series systems under partial repair. In this chapter, we apply the imperfect repair models to the data collected from the signaling systems operated by Bane NOR, aiming to demonstrate the potential benefits in accurately evaluating the system's reliability and its remaining lifetime.

This chapter is organized as follows: a brief description of the signaling system, as well as the collected data, is given in Section 6.1; Section 6.2 presents three case studies in which the imperfect repair models are fitted to the data; after highlighting the necessity of accounting for the heterogeneity, ARA_∞ model, combined with gamma-distributed frailty, is applied to the signal failure data at Brumunddal station in Section 6.3. Concluding remarks are given in Section 6.4.

6.1 General description

Signaling is essentially a sophisticated traffic light system for the railway. The complexities of moving trains around such a large network, keeping them safely apart, and allowing for their long stopping distances, means the signaling system is very complex and comprises a great many parts.

The signals themselves are the line-side pieces of equipment that tell train drivers when it is safe to proceed and what route their train will take. A light signal comprises 1) signal head with light sources, background screen, and shadow screens, 2) mast with platform/ ladder when needed, and 3) devices for controlling the signal with the interface to interlocking equipment.

We focus on the sub-system of the light sources in a signal head, formed by two to five lights of different colors (red, green, yellow, white). One lamp can be either LED or incandescent. A priori, there is no manifest heterogeneity in the lights of the same type, but the maintenance strategy depends on the importance of the lamp: since the failure of

a red light which gives the signal "stop" is often more severe than the failure of a green one, preventive maintenances have been performed only on the signals that consist of red light bulbs. Both corrective and preventive maintenances consist of replacing the light bulb, burned or still working, by a new one. Some preventive maintenances, like periodic inspection or cleaning, are planned for other parts of the signal (cables, covering glass, etc.), and are not considered here.

Since the lamps are not considered to be critical parts, the identity of the failed light bulb is usually not recorded, i.e., we do not know which exact light bulb has been changed. Thus, the failure history of an individual light bulb is not available, resulting in a pooled output. An SRP can, therefore, be used to describe the successive failures of the light sources in a signal.

6.1.1 Data profile

The data we possess contains various information, among which the dates and durations of the recorded maintenances are the most important. As soon as the failure of a signal is spotted (often by the train operator, during maintenance work, or via inspection), maintenance is scheduled and carried out. The delay between the discovery of the failure and the beginning of the repair action is usually not 0, and so is the repair duration. The distribution of CM and PM durations are gathered respectively in Table 6.1 and 6.2.

<= 1 h	<= 24 h	<= 30 days	> 1 months
9240	9566	9715	14

Table 6.1: Repair duration distribution for Corrective maintenance.

<= 24 h	<= 7 days	<= 15 days	<= 1 months	<= 6 months
62036	79629	85281	85749	85806

Table 6.2: Preventive maintenance duration distribution.

Our models are based on point processes, and the repair durations must be neglected. Most recorded CM are completed within one hour and are effectively "negligible" compared to the inter-failure times. The delay between the failure occurrence time and the start of maintenance action is a bit challenging since the former is in practice unknown. Following the suggestion from experts in Bane NOR, we have decided to use the repair start time as the failure time.

6.2 Imperfect repair models fitting

In the following, we investigate particularly the dwarf signals. A dwarf signal is composed of four white lights. Different combinations of lighted bulbs give the train driver signals such as 'Driving prohibited', 'Warning driving allowed', 'Driving permitted', and 'Released for local change'. An error in the light signal that causes light sources not to light must be detectable by interlocking equipment. As we have said, no PM has been implemented

for dwarf signals since there is no red lamp. Figure 6.1 presents a hypothetical trajectory of a maintenance process for a dwarf signal.

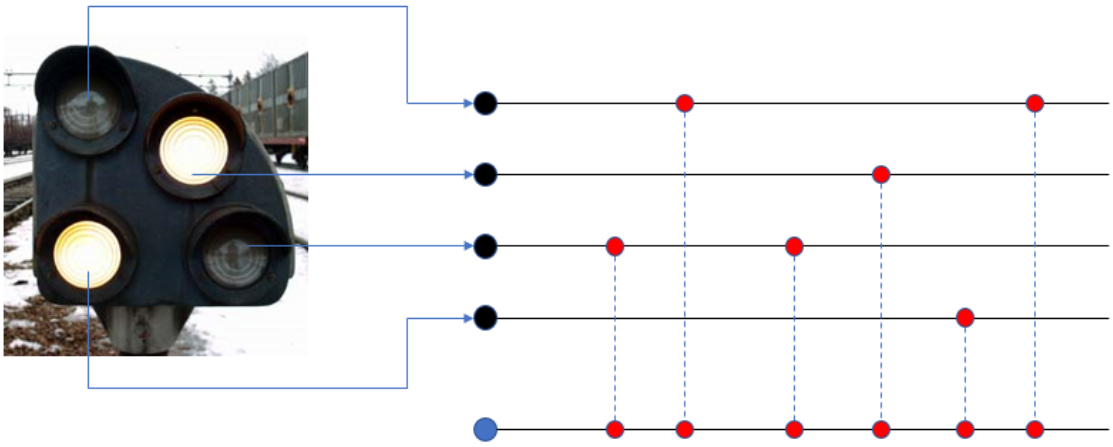


Figure 6.1: Illustrative trajectory of a CM process for a dwarf signal.

Three case studies are presented where the five approximation models proposed in the previous chapter, namely HPP, SIM, IAT_1, ARA_∞, and BP, are fitted to the failures times of dwarf signals. In the first study, a single system is investigated, showing how to use these models to compute the expectation of the remaining lifetime and evaluating the virtual age given the failure history; the second one focuses on a group of frequently failed systems; the third one studies all the dwarf signals, showing a whole picture of the signals' reliability.

	Object	Inference	Censoring	Nb. asset
Case study 1	System 012110	Individual	No	1
Case study 2	Signals having 7^+ failures	Grouped	Yes	16
Case study 3	All dwarf signals	Grouped	Yes	1608

Table 6.3: A summary of the case studies.

The maintenance records provided by Bane NOR contains not only failure histories (CM) but also some invalid (for dwarf signals) PM plans. For the first case study, the observations start at the first recorded CM and end at a failure time, i.e., there is no censored data. For the second and third case studies, the beginning of observation is considered to be the earliest date of the PM plans, determined individually for each system; and the end of observation is the last date of CM, which is 28 March 2019. Thus, for each system, the last inter-event time is right-censored.

6.2.1 Case study 1: System 012110

The investigated signal is located in Dovre Line, a main national connection between Eastern Norway and Trøndelag and further north for passenger and freight traffic. The consecutive lifetimes are shown in Table 6.4.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
133	10	283	763	19	378	203	920

Table 6.4: Inter-failure times of System 012110.

Parameters of fitted models and some reliability indicators are gathered in Table 6.5. The second column A_0 represents the expected virtual age at the beginning of the observation and needs to be calculated only for imperfect maintenance models, namely ARA_∞ and BP, wherein the system is considered to have entered its steady-state and is, therefore, not as-good-as-new at the beginning of observation. The third column $E[X]$ records the unconditional mean lifetime of the system, whereas the last column $E[X_9]$ is the expected value of the 9-th lifetime.

	Parameters	A_0	$E[X]$	$E[X_9]$
ARA_∞	$\alpha = 3.34 \cdot 10^{-10}$ $\beta = 2.9737$ $\rho = 0.1577$	1717.1	321.38	239.83
BP	$\alpha = 3.27 \cdot 10^{-10}$ $\beta = 3.3571$ $p = 0.4113$	460.77	321.91	293.36
HPP	339.08	0	339.08	339.08
IAT_1	$\theta = -1.32$	0	281.52	210.49
SIM		0	281.52	281.52

Table 6.5: Parameters of the models fitted to asset 012110.

For ARA_∞, the parameters (α, β, ρ) are estimated by maximizing the likelihood function defined in Eq.(5.24). The age at the start of the observation, A_0, is assumed to be equal to $E[A_\infty|\alpha, \beta, \rho]$ and is computed using Eq.(2.69). We can then compute the effective ages after each repair by $A_n = (1-\rho)(A_{n-1} + X_n)$. Figure 6.2 shows the variation of the virtual age of the system under consecutive CM. The 9-th lifetime is computed based on A_8, the virtual age after the latest repair. Its survival function is given by

$$R_{X_9}(t|A_8) = e^{-\alpha(t+A_8)^\beta + \alpha A_8^\beta}. \tag{6.1}$$

$E[X_9]$ is obtained by integrating Eq.(6.1). Since the last lifetime X_8 is relatively large, $E[X_9]$ is smaller than the unconditional mean lifetime $E[X]$, which is consistent with the negative correlation between successive intervals in an ARA_∞ process.

As for BP, one should first assess the most likely maintenance effects using the method described in [79]. The repair effect of the 1st-7th maintenance actions can be represented by a binary vector $B = [0, 0, 1, 0, 1, 0, 1]$, where 0 signifies a minimal repair and 1 a perfect one. Remark that the maintenance effect of the last recorded CM is non-identifiable in the absence of further data. The BP parameters are estimated by maximizing the likelihood function defined in Eq.(5.28). A_0 is then calculated using Eq.(2.41), and the consecutive virtual ages are plotted in Figure 6.3. After the 7th repair (just before $X_8 = 920$), the system is assumed to be good as new. The repair after the last observed lifetime has the

probability p to be perfect and $1-p$ to be minimal. Thus, the survival function of the 9th interval is given by

$$R_{X_9}(t|X_8, 7\text{-th repair is perfect}) = pe^{-\alpha t^\beta} + (1-p)e^{-\alpha(t+X_8)^\beta + \alpha X_8^\beta}, \tag{6.2}$$

and $E[X_9]$ under BP model is obtained by integrating the above equation.

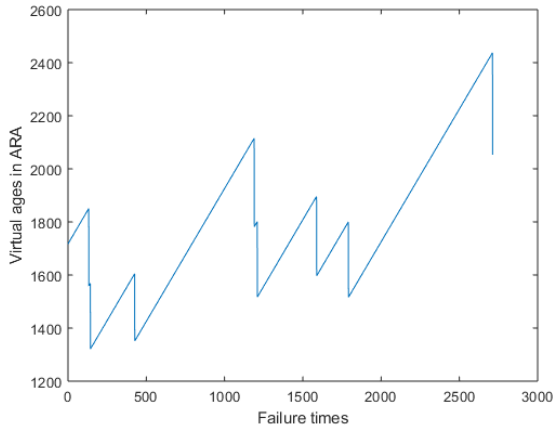


Figure 6.2: Virtual ages in the approximating ARA_∞ model.

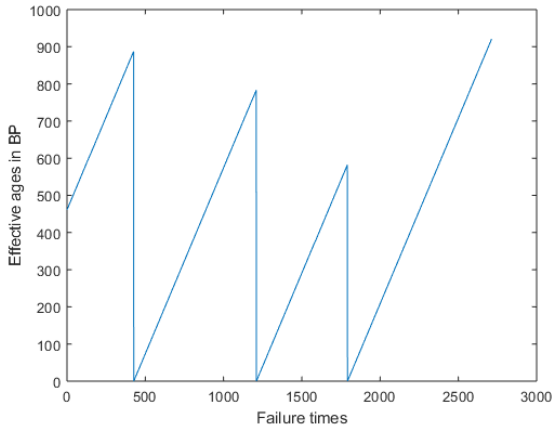


Figure 6.3: Virtual ages in the approximating BP model.

It can be observed from Figures 6.2 and 6.3 that the asset seems to be older under ARA_∞ assumption than it is under the BP model. This is because 1) a low maintenance efficiency was estimated for ARA_∞, which limits the reduction of age at each repair, and 2) three maintenances are estimated to be perfect in BP model, making the asset relatively young. Consequently, the remaining lifetime predicted by ARA_∞ is smaller than that evaluated by BP. Survival functions of the unconditional lifetimes have been plotted in Figure 6.4. It could be noticed that the tails of ARA_∞ and of BP are beneath the exponential fit, which indicates an increasing failure rate when the system has been working long enough.

In addition, ARA_∞ and BP approximations suggest that the system is aging: the Weibull shape β in both models is larger than 1. The IAT_1 does not directly model the failure rate of the system. Instead, it describes the correlation between adjacent intervals with $\theta < 0$. As a result, the expected values of the 9th lifetime calculated under ARA_∞, BP, and IAT_1 are both smaller than $E[X]$, the unconditional mean lifetime.

6.2.2 Case study 2: the most frequently failed systems

One common issue in survival analysis is that failures are generally rare. The power of the statistical inference should be questioned when fitting a parametric model to a small dataset. Therefore, the aggregated data, gathered from a group of similar assets based on the assumption that they are identical, is often used. We investigate the 16 most frequently failed systems and assume that they share the same parameters. The inference procedure for such a group of systems is to maximize the product of the likelihood function of all group members.

The model parameters and reliability indicators are gathered in Table 6.6, and the unconditional survival functions are plotted in Figure 6.5. It can be observed that the exponential curve fits well the Kaplan Meier estimate, signifying a weak dependence between intervals as well as a constant failure rate. The parameter of IAT_1 indicates a weak positive dependence, whereas ARA_∞ and BP suggest an increasing failure rate.

	Parameters	A_0	$E[X]$
ARA_∞	$\alpha = 2.16 \cdot 10^{-12}$ $\beta = 3.5363$ $\rho = 0.2277$	1706.1	502.97
BP	$\alpha = 2.55 \cdot 10^{-7}$ $\beta = 2.2417$ $p = 0.4698$	575.07	509.63
HPP	460.27	0	460.27
IAT_1	$\theta = 0.1365$	0	440.12
SIM		0	440.12

Table 6.6: Reliability indicators for the 16 most frequently failed assets.

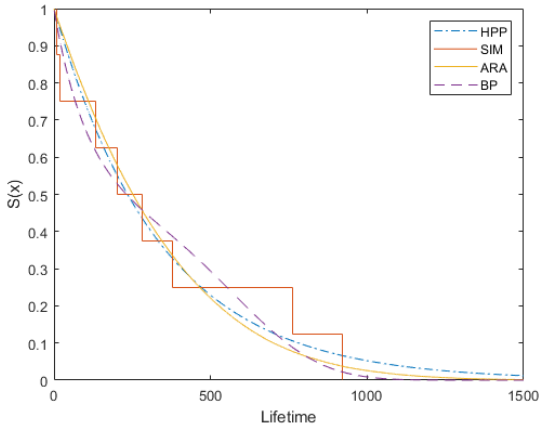


Figure 6.4: Survival functions of the inter-failure times of system 012110.

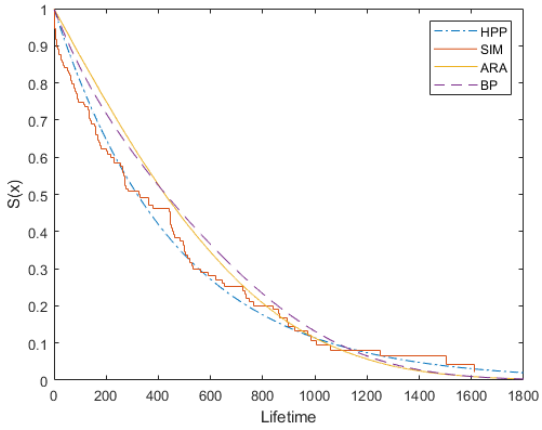


Figure 6.5: Survival function derived from the 16 most frequently failed assets.

The validity of assuming that the most frequently failed system share the same model parameters should be questioned. In fact, it is important to know if the systems under consideration have similar physical characteristics, e.g., assets located at the same station and composed by lamps of the same type may be alike because the external environment is common for each individual. Gathering the systems based on merely the failure numbers lacks a theoretic foundation.

6.2.3 Case study 3: all dwarf signals

1608 assets are investigated, half of which has never had a failure during the observation frame. The following table gathers the distribution of the number of CM:

NB. Failure	0	1	2	3	4	5	6	7	8	9	10	11
NB. Systems	821	419	178	95	51	17	11	8	2	4	0	2

Table 6.7: Distribution of the number of recorded CM.

The data set is, therefore, highly censored. Since there is no proof suggesting that the signals are alike (we know nothing about their external environments or the frequency of usage), we do not attempt to fit imperfect repair models to the aggregated data. Instead, we compare the empirical survival function to the exponential fit, as shown in Figure 6.6.

The fitted exponential model has a mean lifetime 6227 days (around 17 years). As for the Kaplan Meier estimator, if we want to calculate the expected lifetime, we have to change the largest observation to a "death" (as shown in Figure 6.7) since the Kaplan Meier estimator is not defined beyond the largest observation which is right-censored. Integrating the piecewise linear version of the Kaplan-Meier estimator results in a mean lifetime of 2576 days (about 7 years).

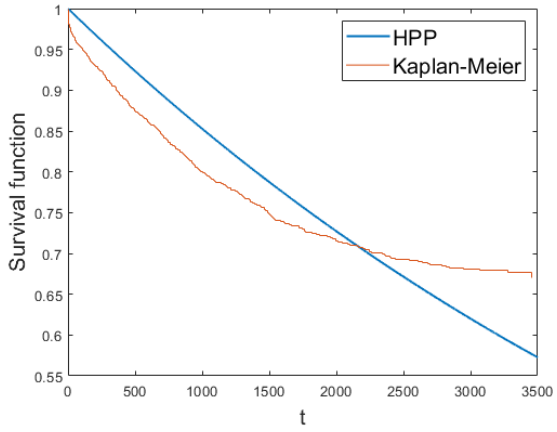


Figure 6.6: Survival functions derived from the failure times of all dwarf signals.

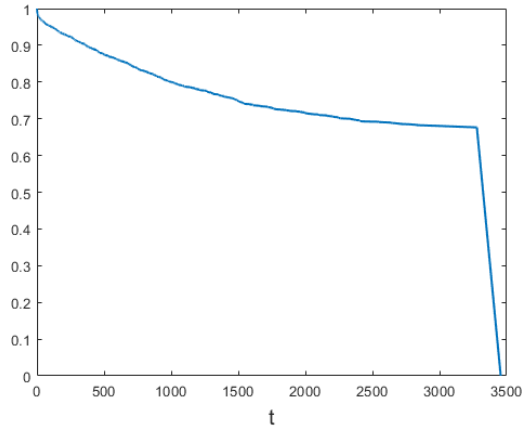


Figure 6.7: Piecewise linear version of the Kaplan-Meier estimator.

The fact that the empirical survival curve crosses its exponential fit once only, and from below, strongly suggests that the failure distribution has a decreasing failure rate [93]. As discussed in Chapter 4, the DFR is possibly related to a heterogeneous population. In our case, the explicit heterogeneity is perhaps related to some measurable covariates, i.e., working environment (average temperature, humidity, precipitation) and usage (total lighted time of the lamps or the frequency of switching on and off), whereas the "unobserved" heterogeneity includes the variation in the quality or robustness of the lamps. For instance, some of the light bulbs are so robust that they can survive the fluctuation of voltage that could kill the others. Without external damage, their lifetime can be up to several decades.

6.3 Heterogeneity

In the following case study, we reveal the influences of unobserved heterogeneity on the evaluation of the system's reliability. The signals at the Brumunddal station are investi-

gated. Their inter-failure times are listed in table 6.8. Previously, five models have been fitted to the SRP dataset, yet, we investigate here only the ARA_∞ approximation: as shown in Chapter 5, Figures 5.1, 5.6, and 5.8, when using ARA_∞ to approximate an SRP, the distance between the density functions is relatively moderate in comparison with BP or HPP. As for the IAT_1 that utilizes a Frank copula to depict the correlation between adjacent intervals, its parameter θ is non-identifiable if no failure has ever occurred during the observation window.

Although the geometrical closeness of the nine signals at Brumunddal station could eliminate certain sources of heterogeneity, such as environmental conditions (humidity, temperature, precipitation, etc.), different frequencies of use (numbers of switching on and off) and manufacturers still constitute an origin of heterogeneity. Since there is currently no available data regarding the relevant covariates, we assume that the heterogeneity lies in gamma-distributed frailty, which makes the hazard rates proportional to each other. This being, each signal would have the same aging rate β (because of similar external environmental conditions) and the same repair effectiveness ρ (performed by the same maintenance crew) and different pseudo scale parameter α.

Systems	Inter-failure times
1	250, 23, 163, 533, 55, 1528, 637*
2	529, 17, 735, 370, 1026, 883*
3	82, 872, 928, 1308*
4	1608, 94, 1488*
5	424, 1887, 879*
6	84, 3106*
7	268, 2922*
8	2667, 523*
9	3190*

Table 6.8: Signal failure times: right censoring is marked with $*$.

6.3.1 Homogeneous assumption

As a comparison, let us consider first fitting an ARA_∞ model to the observations without accounting for the heterogeneity. Since the systems are already in their steady states, the likelihood for a single ARA_∞ sequence, $\mathcal{X}$, can be obtained as indicated in [28]:

$$\tilde{\mathcal{L}}_f^s(\alpha,\beta,\rho|\mathcal{X}) = \int_{(1-\rho)X_1}^{\infty} \mathcal{L}_f^s(\alpha,\beta,\rho|\mathcal{X}, a_0 = \frac{y}{1-\rho} - X_1) f_{A_\infty}(y)dy. \tag{6.3}$$

where $\mathcal{L}_f^s$ is given in Eq.(4.8). The log-likelihood of ARA_∞ population, i.e., the nine systems under consideration, is computed by summing up individual log-likelihoods. The estimated parameters, as well as the 90% confidence intervals (CI, calculated using the observed Fisher information [36]) when the heterogeneity between the systems is ignored,

are therefore given below:

$$\begin{cases} \hat{\alpha} = 1.1833 \cdot 10^{-5}, \quad CI = [1.1828 \cdot 10^{-5}, 1.1838 \cdot 10^{-5}], \\ \hat{\beta} = 1.46, \quad CI = [1.22, 1.70], \\ \hat{\rho} = 0.59, \quad CI = [0.56, 0.64]. \end{cases} \tag{6.4}$$

6.3.2 Heterogeneous assumption

Assume now that the nine systems at Brumunddal station are heterogeneous. With a gamma-distributed frailty, the ARA_∞ model is determined by the quadruple (k, θ, β, ρ) where k and θ are parameters of the gamma distribution followed by α, as described in Eq.(4.2). In Chapter 4, we discussed the inference procedure that allows us to estimate first β and ρ (Eq.(4.19)), then k and θ (Eq.(4.20)) based on an observation matrix $\mathcal{X}_M$. Nevertheless, these two equations are only valid if the virtual age of the systems is known at the beginning of the observation. In the current study, the systems were already in their steady states, and the initial virtual age should be considered as a random variable, i.e., A^p_∞, defined as the virtual age after a repair of an item randomly drawn from the heterogeneous population that has entered the steady-state. Its survival function is given in Eq.(4.3). Consequently, a modification of Eq.(4.19) can be expressed as follows:

$$(\beta^*, \rho^*) = \arg\max_{\beta,\rho} \prod_{j=1}^{M} \tilde{\mathcal{L}}^s_f(\alpha_j(\beta, \rho), \beta, \rho | \mathcal{X}_j), \tag{6.5}$$

with

$$\alpha_j(\beta, \rho) = \arg\max_{\alpha} \tilde{\mathcal{L}}^s_f(\alpha | \beta, \rho, \mathcal{X}_j). \tag{6.6}$$

Using Eqs.(6.6) and (6.5), β and ρ are estimated as

$$\begin{cases} \beta^* = 2.74, \quad CI = [2.56, 2.93], \\ \rho^* = 0.22, \quad CI = [0.03, 0.40]. \end{cases} \tag{6.7}$$

Clearly, when the heterogeneity is overlooked, the estimated aging parameter $\hat{\beta}$ is significantly smaller than β^* and $\hat{\rho}$ is much larger than ρ^*. This is consistent with what has been discovered in Chapter 4. Let us now look into the variation of the pseudo-scale parameter α.

Since the system is in the steady-state at the beginning of the observation, k and θ can no longer be estimated using Eq.(4.23). Instead, they can be determined numerically by maximizing the corresponding likelihood. The likelihood of a single ARA_∞ with an unknown initial age and gamma-distributed α is given by:

$$\tilde{\mathcal{L}}^s_r(k, \theta, \beta, \rho | \mathcal{X}) = \int_{(1-\rho)X_1}^{\infty} \mathcal{L}^s_r(k, \theta, \beta, \rho | \mathcal{X}, a_0 = \frac{y}{1-\rho} - X_1) f^p_{A_\infty}(y) dy, \tag{6.8}$$

where $\mathcal{L}^s_r$ is given by Eq.(4.14) and $f^p_{A_\infty}(t)$ is the Pdf of A^p_∞, obtained by taking the

derivative of Eq.(4.3):

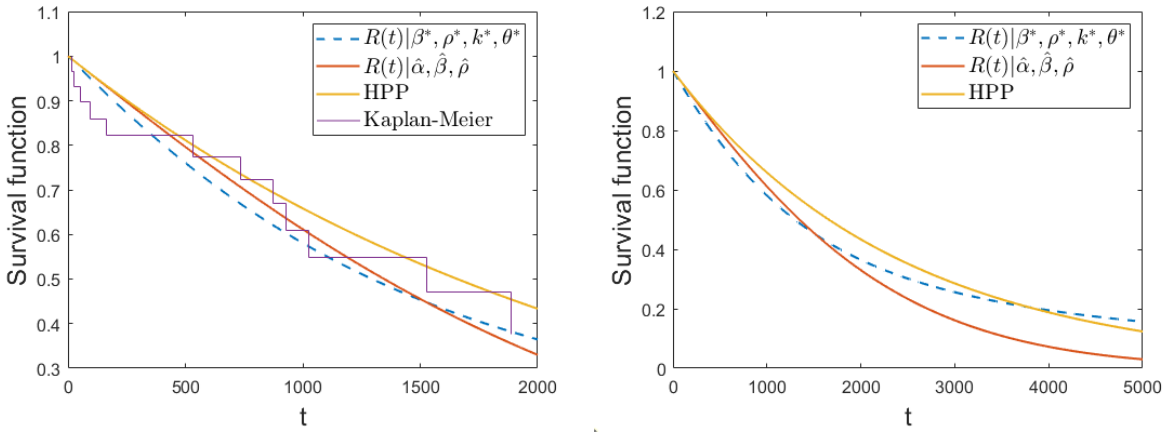
$$f^{p}_{A_\infty}(t) = \sum_{s=1}^{\infty} \frac{1}{(q,q)_\infty (\frac{1}{q},\frac{1}{q})_{s-1}} (1 + \frac{\theta t^\beta}{q^s})^{-k-1} \cdot \frac{k\theta\beta}{q^s} \cdot t^{\beta-1}. \tag{6.9}$$

Correspondingly, the likelihood of the ARA_∞ population, with an unknown initial age and a gamma-distributed α value, given the observation matrix $\mathcal{X}_M$, is the product of individual likelihoods. Having estimated β^* and ρ^*, we can derive k and θ as follows:

$$(k^*, \theta^*) = \arg\max_{k,\theta} \prod_{j=1}^{M} \tilde{\mathcal{L}}^s_r(k, \theta | \mathcal{X}_j, \beta^*, \rho^*). \tag{6.10}$$

Using the equations listed above along with β^* and ρ^*, k and θ are estimated as

$$\begin{cases} k^* = 0.37, \quad CI = [0.18, 0.77], \\ \theta^* = 4.37 \cdot 10^{-10}, \quad CI = [1.18 \cdot 10^{-10}, 1.62 \cdot 10^{-9}]. \end{cases} \tag{6.11}$$



(a) On the scale 1:2000 (b) Tails of ARA_∞ and HPP

Figure 6.8: Kaplan-Meier estimator, HPP and ARA_∞ fit.

We can now compare the survival functions of the time between failure for the signals. Without considering the heterogeneity, the survival function is given by Eq.(2.67) with $\hat{\alpha} = 1.18 \cdot 10^{-5}$, $\hat{\beta} = 1.46$, and $\hat{\rho} = 0.59$ (red solid line in Figure 6.8b). Assuming that the gamma-distributed frailty is present among the signals, the survival function is computed using Eq.(4.4) with $\beta^* = 2.74$, $\rho^* = 0.22$, $k^* = 0.37$ and $\theta^* = 4.37 \cdot 10^{-10}$ (blue dashed line in Figure 6.8b). The pseudo-scale parameter is now regarded as a gamma-distributed random variable with mean $k^*\theta^* = 1.6169 \cdot 10^{-10}$ and variance $k^*\theta^{*2} = 7.0659 \cdot 10^{-20}$. Obviously, the survival function has a heavier tail when the heterogeneity is taken into consideration.

We can further decompose the observed inter-failure times and fit an HPP to the individual sequences while assuming that they share the same failure rate. For HPP, the first intervals in Table 6.8 are considered to be right-censored, i.e., the first lifetime is larger than the observed first interval. The HPP fit is given by the orange line in Figure 6.8, with a mean

lifetime of 2392 days. On the scale of 1:2000, the HPP fit is slightly above the ARA_∞ curves (Figure 6.8a), whereas the tail for HPP is beneath that of the heterogeneous ARA_∞ model as is shown in Figure 6.8b. This agreed with the observed decreasing failure rate when we investigated all the dwarf signals in the previous case study.

6.4 Conclusion

In this chapter, we have investigated the dwarf signals in the Norwegian railway. The signals heads consist of light bulbs in a series configuration and can be, therefore, modeled by the superposition of renewal processes. To study the reliability and remaining lifetime of such system, the most convenient model is HPP, which assumes that the systems have an exponentially distributed inter-failure time. However, the empirical survival function crosses the exponential fit from below when investigating overall the 1608 dwarf signals: the signal head has seemingly a decreasing failure rate.

The DFR, however, is not necessarily due to the DFR of the components that compose the signal head, i.e., the light bulbs. In fact, according to the frailty analysis, the DFR is a result of "weakest objects die out first" reasoning: some light bulbs are more robust than others. This is supported by the fact that more than half of the dwarf signals (821 out of 1608) have never had a failure during ten years of observation.

Each signal head being modeled by an independent SRP, it is not a good idea to decompose the inter-failure times and merge them. That is why we did not try to fit a Weibull renewal process to the inter-failure times of all the systems. Actually, it is essential to treat individual failure sequences as a whole. Otherwise, the intrinsic correlation between adjacent intervals in an SRP will be lost.

Thus, the heterogeneous ARA_∞ seems to be a plausible model that could be used to evaluate the mean lifetime and reliability of the systems: the variation of the pseudo scale α makes it possible to depict a DFR situation even with an increasing baseline failure rate. The sources of heterogeneity for the signal heads include not only the reliability of light bulbs but also the external environment or maintenance crews. Using a heterogeneous ARA_∞ implies that all heterogeneities, no matter the sources, are modeled by the proportional hazard rates.

Finally, for future work, it would be of value to incorporate the fore-mentioned covariates into the current model. This will certainly lead to a better understanding of the aging process by explaining why some systems are more robust than others and will enable the formulation of maintenance plans that are more targeted and cost-effective.

Chapter 7

Conclusion and perspective

Imperfect maintenance models are simple, flexible and realistic. We have first investigated the mathematical properties of some existing models. For those used to describe the situations where failures arrive asymptotically at a constant rate, i.e., having a steady-state, we have proven that they share similar properties as the renewal process: the explicit formulas for the distribution of age, remaining lifetime and spread in a renewal process are also valid for these steady-state imperfect repair models. Moreover, for two virtual age processes, ARA_∞ and BP, we have found the asymptotic distribution of virtual age.

Next, we have studied the mixture of heterogeneous imperfect repair models—an issue that often rises when estimating parameters from aggregated data. For ARA_1, ARA_∞, BP, and the geometric process, we have shown that the unobserved heterogeneity, if overlooked, would result in biased estimates, i.e., an overestimated repair efficiency and an underestimated aging rate. Traditional MLE methods reveal to be inconsistent facing heterogeneity, and alternative estimators have been established. Moreover, parametric models have been proposed for ARA_∞ and BP to characterize the heterogeneity originated in proportional failure rates, aging speed, and repair efficiency.

Then, we postulate the approximation of superposition of renewal processes—a common modeling approach for repairable series systems—by imperfect repair models that generally have less parameters than the SRP and thus allow us to avoid cumbersome inference procedures. We have investigated the approximation errors in the distributions and mean value of lifetimes. Besides, the correlation between adjacent intervals in ARA_∞ and IAT_1 has been proven to be a plausible approximation of the dependence structure between two intervals in an SRP. This enables a better evaluation of the system's aging and a more accurate prognosis.

Finally, the theoretical developments are tested on simulated data as well as data collected from the Norwegian railway network. Fitting imperfect repair models to the signal failure data that was originally modeled by an SRP while accounting for the unobserved heterogeneity, we have highlighted the value of imperfect repair models in evaluating the reliability of systems, assessing the effectiveness of repairs and making optimal maintenance plans.

Based on the current work, several straightforward topics for future research include the

asymptotic properties for imperfect repair models combined with preventive maintenance, the interaction of different sources of heterogeneity and their influence, an explicit formulation of heterogeneity using covariates, to name a few.

From a broader perspective, we believe it is of value to integrate imperfect maintenance models with other mathematical tools such as the Markov chain. Currently, we have only considered binary systems. Allowing the multiple states while adopting the imperfect repair assumptions would possibly lead to a generalization of the Markov renewal process. Another example consists of a variant external environment: ordinary ARA_∞ and BP are governed by a baseline failure intensity, which would become less realistic if the external environment is changing. The asymptotic properties, e.g., the existence of steady-state and limiting the mean lifetime distribution, will also change accordingly under such assumptions. The combination of imperfect maintenance models with shock models, load-sharing systems, or competing risk models are also worth of careful attention in the future research topics.

Appendices

Appendix A

Introduction

A.1 Contexte

De nos jours, de nombreux systèmes, tels que les lignes de production, les équipements d'armes, les centrales nucléaires, les appareils, les véhicules, les avions, etc., sont devenus de plus en plus complexes. Les coûts de leur utilisation deviennent également encore plus élevés qu'auparavant. La maintenance doit être effectuée afin de maintenir les performances de ces systèmes proches du niveau de la conception d'origine. La plupart des systèmes utilisés dans la pratique sont sujets à une détérioration avec l'utilisation et l'âge. Pour ces systèmes détériorés, la maintenance, comme la surveillance, les réparations et les remplacements, peut prolonger leur durée de vie, maintenir la qualité des opérations, réduire le coût des opérations et prévenir les défaillances du système. Notez que, dans ce travail, nous nous intéressons uniquement aux systèmes réparables qui peuvent être restaurés à l'état de fonctionnement via le remplacement ou la réparation de composants en cas de défaillance du système.

En pratique, la maintenance est divisée en deux grandes classes: l'une est dite maintenance corrective (CM), l'autre maintenance préventive (PM). La maintenance corrective vise à restaurer le système dans un état spécifié en cas de défaillance du système. La maintenance préventive vise à maintenir le système dans un état spécifié lorsque le système fonctionne. Avec la tendance actuelle de l'Industrie 4.0, de plus en plus d'efforts sont consacrés aux maintenances préventives, voire proactives. L'objectif est de minimiser la perte potentielle d'une défaillance du système en effectuant des tâches préventives au bon moment. Non seulement devons-nous rassembler toutes sortes de données sur l'état du système, par exemple, l'âge et les caractéristiques physiques, qui nous aident à évaluer la durée de vie restante et l'intensité de défaillance instantanée, mais également un modèle de maintenance approprié est nécessaire pour représenter l'interaction entre les activités de maintenance et correctement le système physique.

Les modèles de maintenance sont à la base de toute analyse de maintenance quantitative, qui peut être utilisé pour analyser et évaluer les performances des approches de maintenance. Plusieurs facteurs distinguent ces modèles de maintenance, notamment, mais sans s'y limiter, les politiques de maintenance (remplacement d'âge, remplacement de bloc, limite de défaillance ...), les structures de système (structure série, structure parallèle, k-

out-of-n ...), degré de maintenance (réparation parfaite, réparation minimale, réparation imparfaite ...), critères d'optimisation (taux de coût minimal, disponibilité maximale, fiabilité maximale ...), distributions de durée de vie (Exponentiel, Weibull, Gamma ...).

Le document actuel se concentre sur les modèles de maintenance imparfaits (IMM) ou les modèles de réparation imparfaits, dans lesquels les activités de maintenance ne doivent pas être parfaites (restaurer le système dans un état neuf) ou minimes (ramener le système dans l'état juste avant l'échec sans aucune amélioration). Bien que le mot «imparfait» indique généralement qu'un système est ramené à un état entre bon comme neuf et mauvais comme vieux, ces modèles sont également capables de modéliser la maintenance nuisible (pire que l'ancien) ou maintenance parfaite (meilleure que neuve) Au cours des dernières décennies, les IMM ont été largement étudiés car ils sont suffisamment flexibles pour représenter une grande échelle de situations, ce qui conduit à son tour à des stratégies de maintenance plus efficaces pour minimiser le taux de coût ou maximiser la fiabilité / disponibilité des actifs. Nous énumérons ici certains des principaux facteurs qui distinguent les modèles de réparation imparfaits:

- Type d'entretien. CM et PM peuvent être imparfaits. Les IMM peuvent impliquer uniquement CM ou CM et PM.
- Mécanisme de réparation. La maintenance est généralement censée réduire l'âge d'un élément, mais parfois, elle est supposée réduire l'intensité de la défaillance.
- Stabilité. Certains modèles sont utilisés pour représenter le système en détérioration avec des pannes se produisant de plus en plus souvent, tandis que d'autres sont utilisés pour modéliser des systèmes ayant un état stationnaire.
- Indépendance. Dans certains modèles, les temps de défaillance consécutifs inter sont supposés indépendants; et dans d'autres modèles comme le processus Brown-Proschan, les temps d'inter-échec sont corrélés.
- Homogénéité des réparations. La réparation imparfaite est souvent caractérisée par un degré de réparation, qui détermine dans quelle mesure un système est restauré. Cela pourrait être une constante ou changer avec le temps.

L'outil mathématique pour les IMM est le processus stochastique: ce n'est pas une surprise car le premier IMM (et toujours très populaire), proposé par Kijima dans les années 1980, porte le nom de «processus de renouvellement généralisé». beaucoup a été fait concernant les propriétés mathématiques de ces processus: nous pouvons prouver que certains modèles sont stables tandis que d'autres ne le sont pas; nous pouvons estimer les paramètres du modèle en fonction des données sur le processus de défaillance/réparation; il existe des tests statistiques nous indiquant si un système peut ou peut ne pas être décrit par un certain modèle compte tenu des temps d'inter échec observés; en attendant, de nombreux articles étudient l'application de ces modèles, par exemple, trouver la politique optimale de gestion des particules lorsque ni CM ni PM n'est parfait. Cependant, certains sujets restent inaperçus, et il faut répondre à quelques questions apparemment simples.

A.2 Objectifs

D'un point de vue mathématique, peut-on en savoir plus sur les IMM? Par exemple, dans un processus de renouvellement, la distribution du temps de récurrence avant/arrière est connue depuis longtemps. Pourrait-il être généralisé à des IMM stables? De plus, si oui, comment influence-t-il les politiques de maintenance?

Les IMM sont souvent appliqués à un système unitaire ou à une population hétérogène composée de systèmes indépendants et identiques. Si l'hétérogénéité est présente dans la population, c'est-à-dire que les systèmes sont similaires mais pas identiques, peut-on encore estimer correctement les paramètres du modèle à partir des observations? Comment tenir compte de l'hétérogénéité et comment influence-t-elle les stratégies de gestion des particules?

Les systèmes en série sont couramment utilisés dans l'ingénierie de la fiabilité. Le système tombe en panne lorsque l'un de ses composants tombe en panne. Si le composant défectueux est remplacé alors que d'autres ne sont pas entretenus, une partie du système peut être considérée comme «renouvelée», ce qui fait du remplacement une réparation imparfaite au niveau du système. Pouvons-nous utiliser quelques IMMs simples pour approximer les systèmes en série? gains et pertes potentiels?

D'un point de vue pratique, un responsable de la maintenance doit généralement choisir un modèle de maintenance approprié parmi plusieurs candidats. Tous représentent le système physique de manière plausible, mais peuvent avoir des résultats différents lors de la prévision de la durée de vie restante et peuvent suggérer différentes stratégies de maintenance préventive. Comment identifier le meilleur (ou le pire) IMM? Comment évaluons-nous leurs performances?

Cette thèse vise à approfondir la découverte des IMM à la fois en théorie et en pratique. Pour être plus précis, les principaux objectifs sont les suivants:

- Étudiez les propriétés mathématiques des IMM stables.
- Modéliser l'hétérogénéité et évaluer son influence sur les politiques de maintenance.
- Proposer de nouveaux modèles pour rapprocher les systèmes en série.
- Évaluer et comparer les performances des modèles proposés.

A.3 Méthodes de recherche

La recherche est à la fois fondamentale et appliquée, exploratoire et explicative. De nouveaux modèles sont développés en combinant les connaissances dans différents domaines, par exemple, l'analyse de survie et l'analyse de fragilité parce que les théories existantes ne sont pas entièrement capables d'expliquer les observations. Ces modèles sont ensuite testés sur des données de terrain, dans le but d'illustrer les avantages potentiels de leur utilisation.

Les propriétés mathématiques et statistiques existantes des IMM, par exemple, les distributions, les corrélations, les convergences, etc., forment la base de nos nouveaux théorèmes et propositions: tous sont basés sur des preuves mathématiques strictes.

Néanmoins, certaines quantités, par exemple, la distribution asymptotique du temps entre défaillances d'un IMM spécifique ou le taux de coût à long terme des activités de réparation lorsque nous voulons évaluer et comparer certaines stratégies de MP, peuvent difficilement être exprimées par des formules explicites. Lorsque cela se produit, nous avons recours à la simulation de Monte Carlo, comme expliqué ci-dessous.

Seule une analyse quantitative est utilisée lors de ce travail. Les données simulées sur le processus de défaillance / réparation d'un système sont collectées via la simulation Monte Carlo, avec un modèle spécifié et des paramètres connus. Ces données incluent généralement les temps d'inter échec (un nombre positif), le type de durée de vie (censuré ou non censuré), le type de réparation (CM ou PM), le degré de réparation (une proportion), etc., d'un système décrit par un certain IMM. Ils ont été utilisés aux fins suivantes:

1. Vérifiez les propositions et les théorèmes mathématiques. Lorsqu'une formule est proposée, elle est d'abord vérifiée avec des données simulées, ce qui permet d'éliminer rapidement les suppositions et les formules erronées. Une fois la preuve terminée, la simulation de Monte Carlo est à nouveau utilisée pour examiner la validité des propositions.

2. Vérifier les procédures d'estimation et évaluer la cohérence et l'efficacité des estimateurs. L'inférence statistique est l'un des principaux problèmes de la recherche: lorsqu'un modèle est adapté aux observations, les paramètres sous-jacents doivent être déduits par une certaine procédure d'estimation, par exemple, l'appariement des moments, l'estimation du maximum de vraisemblance (MLE). La validité des procédures d'inférence est examinée en comparant les estimations aux paramètres connus. De plus, le réglage de la taille des données permet d'évaluer la cohérence et l'efficacité des estimateurs.

3. Évaluer les stratégies d'entretien quantitativement. Nous avons choisi le taux de coût de réparation à long terme comme critère d'optimisation, qui implique souvent la fonction de renouvellement, c'est-à-dire le nombre prévu de défaillances dans une période de temps. Il s'agit d'une quantité typique qui ne possède pas de formule explicite. Simuler le processus de défaillance/réparation d'un grand nombre de systèmes indépendants et enregistrer les coûts de réparation nous permet d'obtenir empiriquement le taux de coût de réparation à long terme.

Les données de terrain de Bane NOR ont également été analysées. Les enregistrements de défaillance/réparation du système de signalisation ferroviaire norvégien au cours de la dernière décennie contiennent des informations telles que la date d'intervention, la durée de la réparation, le type de réparation, ainsi que l'emplacement des signaux, le fabricant et l'administration. Les données sont ensuite utilisées pour découvrir la carence des IMM existants et pour tester et évaluer les nouveaux modèles proposés.

A.4 Portée et limites de la recherche

Ces travaux portent principalement sur les IMM et l'optimisation quantitative de la maintenance. Ainsi, de nombreux problèmes et techniques de modélisation concernant la fiabilité et l'analyse des risques ne sont pas traités. Ci-dessous, nous énumérons certaines des principales limitations dans le cadre d'un modèle de réparation imparfait:

- Une représentation binaire du système est utilisée: ils peuvent fonctionner ou échouer. Par conséquent, la modélisation multi-états et les approches connexe, par exemple, les chaînes de Markov, ne sont pas poursuivies. Dans une telle hypothèse binaire, les modèles étudiés dans cet article ne peuvent pas être appliqués à des systèmes qui subissent une dégradation continue, tels que le processus de corrosion des tuyaux, la longueur de fissure de la chaussée, la déviation géométrique des voies ferrées.
- Le temps de réparation est considéré comme négligeable. Il s'agit d'une hypothèse quintessentielle en ingénierie de la fiabilité, visant à simplifier le modèle lorsque le temps de réparation est négligeable par rapport à son temps moyen entre défaillances. Bien que cette hypothèse soit valide pour les données fournies par Bane NOR dans le sens où les signaux ferroviaires survivent généralement des années, voire des décennies, alors que l'activité de réparation ne dure généralement pas plus d'une journée, elle limite la gamme d'applications de nos résultats.

Du point de vue de l'optimisation de la maintenance, cette étude présente les limites suivantes:

- L'optimisation de la maintenance que nous abordons dans cette étude n'est qu'une des nombreuses étapes d'une gestion de maintenance réelle. C'est un problème de décision dans la situation "idéale". L'application d'un tel plan de maintenance "optimal" n'est généralement pas aisée en pratique. Un problème typique pourrait être un arriéré élevé, ce qui entrave la mise en œuvre de maintenances préventives.
- Les covariables ne sont pas prises en compte dans cette étude. C'est-à-dire que tous les indicateurs de fiabilité d'un actif, par exemple la durée de vie moyenne, la fréquence de défaillance, la vitesse de vieillissement, sont dérivés uniquement des données de durée de vie sans utiliser d'autres mesures qui pourraient éventuellement aider à évaluer l'état actuel d'un actif et son durée de vie restante. Par exemple, dans les données fournies par Bane NOR, les dates de défaillance/réparation des lampes de signalisation sont enregistrées, ce qui nous indique leur durée de vie et leur durée de réparation. Nous pouvons mieux comprendre leur mécanisme de défaillance et planifier la PM en conséquence si nous connaissons l'heure de leur mise sous/hors tension, qui est logiquement une variable pertinente pour les ampoules.
- Nous déployons une méthode purement quantitative, qui repose fortement sur la qualité et la taille des données collectées. Fondamentalement, nous dérivons la vitesse de vieillissement des actifs des enregistrements de défaillance/réparation, et sur la base des paramètres estimés du modèle, une stratégie PM est proposée. Par conséquent, la puissance statistique lorsque la taille des données est petite doit être remise en question. De plus, ce n'est pas un cadre qui peut facilement intégrer les opinions d'experts.

Pour résumer, les limites mentionnées ci-dessus indiquent certaines orientations de recherche futures: combiner les IMM avec un système de dégradation d'état multi-états ou continu, où la durée de réparation ou le temps de séjour dans différents états n'est plus négligeable; incorporer des covariables dans les modèles; développer éventuellement un cadre bayésien pour prendre en compte l'avis des experts.

Le reste du rapport comprend quatre chapitres. Dans le chapitre 2, nous présentons l'état de l'art sur les IMM. Certains des modèles les plus populaires sont présentés de manière détaillée, avec des définitions mathématiques appropriées. Le chapitre 3 est consacré aux IMM stables. Nous prouvons qu'ils ont des propriétés mathématiques similaires à celles du processus de renouvellement, avant de révéler les distributions asymptotiques des quantités les plus cruciales, y compris l'âge virtuel, dans les IMM. Dans le chapitre 4, nous discutons de l'hétérogénéité des IMM. L'objet étudié n'est plus un système unique, mais un groupe de systèmes similaires qui subissent des réparations imparfaites. Le chapitre 5 est consacré aux systèmes en série, à la superposition des processus de renouvellement et à ses approximations. Nous étudions la relation interne entre le système série et les IMM et proposons de nouvelles approches d'approximation qui aident à prédire la durée de vie restante des systèmes série. Au chapitre 6, les modèles étudiés sont testés sur des données simulées et des données de terrain. En particulier, nous abordons le problème de l'optimisation de la maintenance en utilisant les conclusions des chapitres 3 et 4. Des remarques finales et des discussions sont données au chapitre 7.

Appendix B

Chapter 3

B.1 Introduction

Ce chapitre traite des propriétés asymptotiques pertinentes pour les processus d'âge virtuel à l'état stationnaire. On montre que les distributions limites d'âge, la durée de vie résiduelle et la vie totale qui décrivent un processus de renouvellement ordinaire peuvent être généralisées au processus d'âge virtuel stable, bien que les cycles de ce dernier ne soient pas indépendants. Les distributions asymptotiques de l'âge virtuel au moment t, ainsi que des âges virtuels au début et à la fin d'un cycle contenant t (quand t tend vers l'infini) sont explicitement dérivées pour deux modèles de maintenance imparfaite populaires dans la pratique, à savoir la réduction arithmétique de l'âge à mémoire infinie (ARA_∞) et les modèles Brown-Proschan (BP). Certaines applications des résultats obtenus à l'optimisation de la maintenance sont discutées.

Comme indiqué dans la section précédente, les processus de renouvellement ordinaires sont stationnaires dans le sens où la fonction de densité de renouvellement correspondante a tendance à être constante lorsque le temps tend vers l'infini. Le NHPP qui décrit les réparations minimales est, de toute évidence, non stationnaire, et si, par exemple, son taux augmente, les défaillances arrivent plus fréquemment avec le temps. Le type Kijima I et le processus géométrique ([79]), comme NHPP, ne sont pas stationnaires et peuvent être utilisés pour modéliser la durée de vie avec les tendances. Bien qu'il existe de nombreuses publications sur diverses applications des modèles d'âge virtuel en matière de fiabilité, peu a été fait dans la littérature sur la description des propriétés asymptotiques pertinentes des processus d'âge virtuel correspondants.

Il convient de noter que les propriétés limitantes des processus de renouvellement ordinaires sont particulièrement importantes dans diverses applications. Par exemple, l'obtention des fonctions de renouvellement correspondantes peut être difficile à calculer, et les valeurs asymptotiques simples fournies par les théorèmes de type renouvellement sont avantageuses dans la pratique. Un autre exemple est le processus de renouvellement alterné. La disponibilité stationnaire dans ce cas, qui est généralement du principal intérêt, est obtenue simplement via les temps de montée et de descente moyens d'un système. Les cycles de vie de nombreux systèmes industriels sont assez longs, ce qui signifie qu'un grand nombre d'actions de maintenance sont effectuées. De plus, dans de nombreux cas, les

données opérationnelles ne sont enregistrées que lorsqu'un système entre dans son régime stable. Par conséquent, l'importance des méthodes asymptotiques dans le contexte décrit est difficile à surestimer.

L'étude des propriétés asymptotiques des processus de réparation imparfaits qui décrivent plus adéquatement que les processus de renouvellement ordinaires la maintenance des systèmes du monde réel, semble être *une tâche naturelle et pratiquement saine* qui est abordée dans le présent article. Pour atteindre cet objectif, nous avons d'abord dû répondre aux questions suivantes: les résultats asymptotiques pour l'âge, la durée de vie résiduelle et la vie totale pour les processus de renouvellement ordinaires peuvent-ils être généralisés (et dans quelles conditions) au cas des processus de réparation imparfaits? Quelles sont les distributions asymptotiques de ces quantités? Pour répondre à ces questions, des résultats théoriques spécifiques ont dû être obtenus et illustrés par plusieurs exemples pratiques.

B.2 Conclusion

Ce chapitre étudie les distributions asymptotiques pour les processus d'âge virtuel stable. Nous montrons d'abord que les distributions limites du temps de récurrence en arrière, de la durée de vie restante et de la vie totale qui caractérisent un processus de renouvellement ordinaire peuvent être généralisées au cas des processus d'âge virtuel avec des cycles asymptotiquement répartis de manière identique. Ensuite, nous dérivons de nouvelles expressions analytiques pour toutes les distributions limites d'intérêt. Nous discutons également de l'importance du mécanisme de réduction de l'âge pour les résultats obtenus. Les exemples fournis mettent en évidence la valeur pratique de nos résultats en ingénierie de la fiabilité.

Ce sujet pourrait être poursuivi à l'avenir dans plusieurs directions. Par exemple, des distributions asymptotiques dans des modèles d'âge virtuel stables impliquant des maintenances préventives imparfaites peuvent être envisagées. Un exemple typique est le processus ARA_1 CM-ARA_∞ PM décrit dans [32]: les maintenances correctives du type ARA_1 ne peuvent pas maintenir le système réparé dans un état stable , tandis que la stationnarité peut être obtenue par les PM périodiques de type ARA $_\infty$. Par conséquent, il pourrait être intéressant d'examiner la distribution asymptotique de l'âge virtuel juste après le PM dans ce cas. Limite distributions dans d'autres modèles de maintenance imparfaits tels que le modèle de réduction arithmétique de l'intensité à mémoire infinie (ARI_∞) ([34]) peut également être digne d'être approfondi.

Appendix C

Chapter 4

C.1 Introduction

Ce chapitre étudie l'effet de l'hétérogénéité sur les défaillances des systèmes réparables qui subissent des réparations imparfaites, qui sont largement utilisées en ingénierie de la fiabilité. Lorsque l'on considère un groupe de systèmes similaires, l'hypothèse que les processus de réparation sont indépendants et distribués de manière identique devient discutable en raison de l'hétérogénéité non observée de ces systèmes. Les modèles de base que nous considérons incluent ARA_∞, ARA_1, Brown Proschan et le processus géométrique.

Dans la section 4.1, pour le processus ARA_∞, nous utilisons le modèle de fragilité pour étudier le taux de risque de base proportionnel entre les systèmes et, en particulier, la fragilité charactérisée par la loi gamma est étudiée. Ainsi, nous dérivons les propriétés asymptotiques du processus de réparation mixte et les estimations de vraisemblance correspondantes, puis évaluons les effets sur l'estimation des paramètres du modèle lorsque l'hétérogénéité est ignorée par erreur. De plus, lorsque le modèle est correctement établi en tenant compte de la fragilité, nous constatons que l'estimateur du maximum de vraisemblance est incohérent et nous proposons une approche alternative. Deux études de cas sont présentées pour illustrer les avantages de la prise en compte de l'hétérogénéité non observée dans la planification des activités de maintenance préventive.

Dans la section 4.2, pour le processus BP, non seulement le taux de risque proportionnel est pris en compte, mais de nouveaux modèles paramétriques ont également été proposés pour décrire l'efficacité de réparation hétérogène et l'environnement de travail. L'impact de l'hétérogénéité sur la durée de vie moyenne de la population est révélé, et en particulier, nous soulignons l'importance d'employer une équipe de maintenance toujours compétente.

Dans la section 4.3, pour ARA_1 et le processus géométrique, l'influence du paramètre d'échelle hétérogène a été abordée. Nous nous concentrons spécifiquement sur la question de l'estimation des paramètres. Il est démontré que lorsque l'hétérogénéité est négligée par erreur, le paramètre de vieillissement est constamment sous-estimé.

Enfin, certaines applications des résultats pertinents dans l'analyse du système de signalisation ferroviaire sont présentées au chapitre 6, montrant les avantages de la prise en

compte de l'hétérogénéité lors de l'estimation de la vitesse et de la fiabilité du vieillissement du système.

C.2 Conclusion

Dans ce chapitre, nous avons considéré le cas où l'hétérogénéité entre les systèmes est combinée avec des modèles de réparation imparfaits stables. Pour le processus ARA_∞, après avoir étudié l'influence d'une fragilité non spécifiée sur la durée de vie moyenne de la population, nous avons examiné spécifiquement la distribution gamma sur le paramètre pseudo-échelle du processus ARA_∞ et dérivé le propriétés asymptotiques, y compris les distributions de la durée moyenne du cycle de la population et de la VA de la population, puis présenté les conséquences pour les cas où l'hétérogénéité entre les systèmes a été erronément ignorée. En particulier, lorsque le modèle était mal spécifié, le taux de vieillissement était sous-estimé, tandis que l'efficacité de la réparation était surestimée. De plus, en raison de la spécialité de ARA_∞, c'est-à-dire des intervalles dépendants, le MLE établi sur le modèle correct était toujours incohérent. Par conséquent, une approche alternative a été proposée et sa cohérence a été vérifiée. Enfin, la question de l'optimisation de la maintenance pour la population qui subit une réparation imparfaite est abordée, et les avantages de la prise en compte de l'hétérogénéité lors de la planification des activités de maintenance préventive sont démontrés.

Pour le modèle de Brown Proschan, nous avons étudié non seulement l'hétérogénéité due à la fragilité, mais également pris en compte l'efficacité d'entretien hétérogène. Il a été souligné que pour minimiser le nombre moyen à long terme de défaillances de systèmes, il est préférable d'employer une équipe habile «moyenne» plutôt qu'une combinaison de personnel hautement qualifié et de quelques recrues.

Enfin, pour deux processus de réparation imparfaits et instables, ARA_1 et le processus géométrique, nous montrons que l'estimation des paramètres du modèle sans tenir compte de l'hétérogénéité cachée peut conduire à des estimations biaisées. Plus précisément, le paramètre de vieillissement, lorsque la distribution de référence est Weibull, est toujours sous-estimé. Il s'agit d'un fait important qui souligne l'importance de tenir compte de l'hétérogénéité du système lors de l'analyse de l'historique de maintenance; sinon, avec le paramètre de vieillissement sous-estimé et une efficacité de réparation souvent surestimée, il est difficile d'optimiser le plan de maintenance.

Les applications industrielles de ces résultats seront démontrées au chapitre 6, où une étude de cas sur les données de défaillance collectées à partir de signaux le long d'une gare norvégienne est réalisée, illustrant les avantages de la prise en compte de l'hétérogénéité non observée dans l'estimation de la vitesse et de la fiabilité du vieillissement des actifs et dans la planification des activités de maintenance préventive.

Néanmoins, la présente étude comporte plusieurs limites. D'abord et avant tout, aucune covariable n'est prise en compte. En pratique, les covariables expliquent les sources de l'hétérogénéité et permettent de mesurer/contrôler plus efficacement, alors que dans notre étude, l'hétérogénéité est considérée comme une propriété intrinsèque des systèmes. Cette question sera à nouveau abordée au chapitre 6 lorsque nous analyserons l'historique des défaillances des systèmes de signalisation ferroviaire exploités par Bane NOR. Deuxièmement,

il serait intéressant d'étudier comment différentes hétérogénéités interagissent les unes avec les autres, étant donné que nous n'avons étudié que *séparément* le taux de défaillance hétérogène, l'efficacité de la réparation hétérogène et la vitesse de vieillissement différente. Nous pensons que ces deux directions méritent une enquête plus approfondie.

Appendix D

Chapter 5

D.1 Introduction

La superposition des processus de renouvellement (SRP) consiste en l'observation des temps entre arrivées sur la base de multiples processus de renouvellement indépendants. Dans l'analyse de maintenance imparfaite, SRP caractérise le système en série réparable lorsque la maintenance consiste à remplacer le composant défaillant tout en laissant les autres composants inchangés (réparés de façon minimale). Les composants peuvent être des structures physiques ou virtuelles comme dans une situation de risque concurrentiel. Après une réparation, le système est souvent entre les états aussi bon que neuf et aussi mauvais que vieux, car un seul des composants a été renouvelé.

Dans la pratique, les enregistrements de maintenance sont souvent incomplets et les informations sur l'identité des composants défaillants ne sont pas nécessairement disponibles. Par conséquent, les observations sont généralement réduites à une sortie groupée [21], constituée des temps de défaillance au niveau du système. L'évaluation de la santé du système, comme son vieillissement global et son efficacité de maintenance, commence généralement par l'estimation des paramètres du modèle. Lorsque le nombre de composants est connu et que tous les composants sont identiques, des procédures d'inférence ont été effectuées par [118] directement sur la sortie regroupée à l'aide d'un partitionnement informatique. Néanmoins, lorsque le nombre d'événements regroupés est limité, ou lorsque les composants sont différents les uns des autres, les méthodes d'inférence directe sont difficiles à mettre en œuvre, inspirant diverses approches d'approximation.

Étant donné que les temps d'inter-occurrence d'un SRP tendent vers une distribution d'équilibre, une simplification naturelle consiste à utiliser un processus de renouvellement [69, 113, 105] pour approximer le SRP. Il existe plusieurs choix pour la distribution à vie du processus de renouvellement approximatif. Deux d'entre eux sont les plus courants: le premier est la distribution exponentielle, qui forme un processus de Poisson homogène (HPP) [113]; la seconde est la distribution limite des intervalles dans la SRP approximé, donnée dans l'équation (**??**), ce qui garantit que le processus de renouvellement approximatif et la SRP approximé ont la même durée de vie attendue. Ceci est appelé dans [105] la méthode à intervalle stationnaire (SIM).

Cependant, le principal inconvénient de l'approximation d'un SRP par un processus de renouvellement est la perte de dépendance entre les temps d'inter-occurrence, en particulier deux temps d'inter-arrivée successifs. Dans ce chapitre, nous proposons trois nouvelles approches d'approximation d'un SRP basées sur des modèles d'âge virtuels et une copule, avant de les comparer aux modèles existants, par exemple, HPP et SIM. Puisque nous nous intéressons principalement à la situation où les paramètres SRP sont inconnus, nous présenterons d'abord les cinq approches tout en mettant l'accent sur les procédures d'estimation qui dérivent les paramètres des modèles d'approximation à partir de la sortie groupée observée d'un SRP. Ensuite, nous évaluons les performances des méthodes proposées, c'est-à-dire pour voir si elles peuvent conserver correctement la durée de vie moyenne ou capturer la corrélation entre des intervalles adjacents ou utiliser d'autres méthodes pour accéder à la " distance " entre le SRP et ses approximations.

D.2 Conclusion

Dans ce chapitre, nous avons présenté et comparé cinq modèles approximatifs pour un SRP: deux modèles de réparation imparfaits (ARA_∞ et Brown-Proschan), deux processus de renouvellement (SIM et HPP) et IAT_1, construits par la distribution marginale estimée de Kaplan-Meier et une copule de Frank qui capture la structure de dépendance entre les intervalles successifs. Les performances de ces modèles sont évaluées en étudiant l'amplitude des erreurs de longueur d'intervalle moyenne et les corrélations lorsque les approximations mentionnées ci-dessus sont utilisées. En outre, leurs capacités dans le pronostic de RUL sont examinées en utilisant l'approche level-set et l'indice de Gini. Il est difficile de dire quel modèle est globalement le meilleur: leurs performances dépendent du taux de vieillissement ainsi que de la quantité de données disponibles.

Les avantages de l'utilisation de ces approches d'approximation seront démontrés dans le chapitre suivant, où les données de défaillance collectées à partir du système de signalisation ferroviaire norvégien seront étudiées. Nous montrons comment ces modèles pourraient être utilisés pour évaluer la fiabilité du système et la durée de vie utile restante, et soulignons en particulier le gain de l'utilisation de modèles de réparation imparfaits qui conservent la dépendance négative entre les intervalles dans un SRP.

Appendix E

Chapter 6

E.1 Introduction

Auparavant, nous avons étudié les propriétés asymptotiques et l'hétérogénéité des modèles de réparation imparfaits. Plus précisément, au chapitre 5, nous avons examiné comment ils pourraient se rapprocher de la superposition des processus de renouvellement - une approche de modélisation commune pour les systèmes en série. Dans ce chapitre, nous appliquons les modèles de réparation imparfaits aux données collectées des systèmes de signalisation exploités par Bane NOR, dans le but de démontrer les avantages potentiels de l'évaluation de la fiabilité du système et de la durée de vie restante.

La signalisation est essentiellement un système sophistiqué de feux de circulation pour le chemin de fer. La complexité de déplacer des trains autour d'un si grand réseau, de les maintenir en sécurité et de tenir compte de leurs longues distances d'arrêt, signifie que le système de signalisation est très compliqué et comprend de nombreuses parties.

Les signaux eux-mêmes sont les équipements côté ligne qui indiquent aux conducteurs de train quand il est sûr de continuer et quelle route leur train prendra. Un signal lumineux comprend 1) une tête de signal avec des sources de lumière, un écran de fond et des écrans d'ombre, 2) un mât avec plate-forme/échelle si nécessaire, et 3) des dispositifs pour contrôler le signal avec l'interface avec l'équipement de verrouillage.

Nous nous concentrons sur le sous-système des sources lumineuses dans une tête de signal, formé de deux à cinq lumières de couleurs différentes (rouge, vert, jaune, blanc). Une lampe peut être à LED ou à incandescence. A priori, il n'y a pas d'hétérogénéité manifeste dans les feux du même type, mais la stratégie de maintenance dépend de l'importance de la lampe: car la panne d'un feu rouge qui donne le signal "stop" est souvent plus grave que la panne de vert, des entretiens préventifs ont été effectués uniquement sur les signaux constitués d'ampoules rouges. Les maintenances correctives et préventives consistent à remplacer l'ampoule, brûlée ou en fonctionnement, par une neuve. Certains entretiens préventifs, comme l'inspection périodique ou le nettoyage, sont prévus pour d'autres parties du signal (câbles, verre de couverture, etc.) et ne sont pas pris en compte ici.

Étant donné que les lampes ne sont pas considérées comme des pièces critiques, l'identité

de l'ampoule défectueuse n'est généralement pas enregistrée, c'est-à-dire que nous ne savons pas quelle ampoule exacte a été changée. Ainsi, l'historique des défaillances d'une ampoule individuelle n'est pas disponible, ce qui entraîne une sortie groupée. Un SRP peut donc être utilisé pour décrire les défaillances successives des sources lumineuses dans un signal.

Ce chapitre est organisé comme suit: une brève description des données collectées est donnée dans la section 6.1; La section 6.2 présente trois études de cas dans lesquelles les modèles de réparation imparfaits sont ajustés aux données; après avoir souligné la nécessité de tenir compte de l'hétérogénéité, le modèle ARA_∞, combiné à la fragilité, est appliqué aux données de défaillance du signal à la station de Brumunddal dans la section 6.3. Des remarques finales sont données dans la section 6.4.

E.2 Conclusion

Dans ce chapitre, nous avons étudié les signaux nains dans le chemin de fer norvégien. Les têtes de signaux sont constituées d'ampoules dans une configuration en série et peuvent donc être modélisées par la superposition de processus de renouvellement. Pour étudier la fiabilité et la durée de vie restante d'un tel système, le modèle le plus pratique est HPP, qui suppose que les systèmes ont un temps d'inter-échec distribué de façon exponentielle. Cependant, cela contredit le fait que la fonction de survie empirique croise l'ajustement exponentiel d'en bas lors de l'enquête globale sur les 1608 signaux nains: la tête de signal a apparemment un taux d'échec décroissant.

Le DFR, cependant, n'est pas nécessairement dû au DFR des composants qui composent la tête de signal, c'est-à-dire les ampoules. En fait, selon l'analyse de fragilité, le DFR est le résultat d'un raisonnement "les objets les plus faibles s'éteignent en premier": certaines ampoules sont plus robustes que d'autres. Ceci est soutenu par le fait que plus de la moitié des signaux nains (821 sur 1608) n'ont jamais connu de panne pendant dix ans d'observation.

Chaque tête de signal étant modélisée par un SRP indépendant, ce n'est pas une bonne idée de décomposer les temps d'inter-échec et de les fusionner. C'est pourquoi nous n'avons pas essayé d'adapter un processus de renouvellement de Weibull aux temps d'interruption de tous les systèmes. En fait, il est essentiel de traiter les séquences de défaillance individuelles dans leur ensemble. Sinon, la corrélation intrinsèque entre les intervalles adjacents dans un SRP sera perdue.

Cela étant, l'hétérogène ARA_∞ semble être un modèle plausible qui pourrait être utilisé pour évaluer la durée de vie moyenne et la fiabilité des systèmes: la variation de la pseudo-échelle α permet de représenter un situation DFR même avec un taux d'échec de base croissant. Les sources d'hétérogénéité des têtes de signal incluent non seulement la fiabilité des ampoules mais aussi l'environnement extérieur ou les équipes de maintenance. L'utilisation d'un ARA_∞ hétérogène implique que toute hétérogénéité, quelle que soit sa provenance, soit modélisée par les taux de risque proportionnels.

Enfin, pour les travaux futurs, il serait utile d'incorporer les covariables susmentionnées dans le modèle actuel. Cela permettra certainement de mieux comprendre le processus de

vieillissement en expliquant pourquoi certains systèmes sont plus robustes que d'autres et permettra la formulation de plans de maintenance plus ciblés et plus rentables.

Appendix F

Conclusion

Les modèles de maintenance imparfaits sont populaires dans la pratique en raison de leur simplicité et de leur flexibilité simultanées. Nous avons d'abord étudié les propriétés mathématiques de certains modèles existants. Pour ceux utilisés pour décrire les situations où les échecs arrivent à un taux constant, c'est-à-dire ayant un état stationnaire, nous avons prouvé qu'ils partagent des propriétés similaires au processus de renouvellement: les formules explicites pour la distribution de l'âge, la durée de vie restante et la propagation dans un processus de renouvellement est également valable pour ces modèles de réparation imparfaits en régime permanent. De plus, pour deux processus d'âge virtuel, ARA_∞ et BP, nous avons trouvé la distribution asymptotique de l'âge virtuel.

Ensuite, nous avons étudié le mélange de modèles de réparation hétérogènes imparfaits — un problème qui se pose souvent lors de l'estimation des paramètres à partir de données agrégées. Pour ARA_1, ARA_∞, BP et processus géométrique, nous avons montré que l'hétérogénéité non observée, si elle était négligée, entraînerait des estimations biaisées, c'est-à-dire une efficacité de réparation surestimée et un taux de vieillissement sous-estimé. Les méthodes MLE traditionnelles pourraient être incohérentes face à l'hétérogénéité, et d'autres estimateurs ont été établis. De plus, des modèles paramétriques ont été proposés pour ARA_∞ et BP pour caractériser l'hétérogénéité provenant des taux de défaillance proportionnels, de la vitesse de vieillissement et de l'efficacité des réparations.

Ensuite, nous postulons l'approximation de la superposition des processus de renouvellement — une approche de modélisation commune pour les systèmes en série réparables — par des modèles de réparation imparfaits qui ont généralement moins de paramètres que le SRP et nous permettent ainsi d'éviter des procédures d'inférence lourdes. Nous avons étudié les erreurs d'approximation dans les distributions et la valeur moyenne des durées de vie. En outre, la corrélation entre les intervalles adjacents dans ARA_∞ et IAT_1 s'est avérée être une approximation plausible de la structure de dépendance entre deux intervalles dans un SRP. Cela permet une meilleure évaluation du vieillissement du système et un pronostic plus précis.

Enfin, les développements théoriques sont testés sur des données simulées ainsi que des données collectées sur le réseau ferroviaire norvégien. En adaptant les modèles de réparation imparfaits aux données de défaillance du signal qui ont été initialement modélisées par un SRP tout en tenant compte de l'hétérogénéité non observée, nous avons souligné la

valeur des modèles de réparation imparfaits dans l'évaluation de la fiabilité des systèmes, l'évaluation de l'efficacité des réparations et l'élaboration de plans d'entretien optimaux.

Sur la base des travaux en cours, plusieurs sujets simples pour de futures recherches comprennent les propriétés asymptotiques pour les modèles de réparation imparfaits combinés à la maintenance préventive, l'interaction de différentes sources d'hétérogénéité et leur influence, une formulation explicite d'hétérogénéité à l'aide de covariables, pour n'en nommer que quelques-unes.

Dans une perspective plus large, nous pensons qu'il est utile d'intégrer des modèles de maintenance imparfaits à d'autres outils mathématiques tels que la chaîne de Markov. Actuellement, l'état du système est binaire. Autoriser les états multiples tout en adoptant les hypothèses de réparation imparfaites pourrait éventuellement conduire à une généralisation du processus de renouvellement de Markov. Un autre exemple consiste en une variante d'environnement externe: les ARA_∞ et BP ordinaires sont régis par une intensité de défaillance de base, qui deviendrait moins réaliste si l'environnement externe change. Les propriétés asymptotiques, par exemple, l'existence d'un état stationnaire et la limitation de la distribution de durée de vie moyenne, changeront également en conséquence sous de telles hypothèses. La combinaison de modèles de maintenance imparfaits avec des modèles de choc, des systèmes de partage de charge ou des modèles de risque concurrents mérite également quelques recherches.

Bibliography

[1] Odd O. Aalen. Modelling heterogeneity in survival analysis by the compound poisson distribution. *Ann. Appl. Probab.*, 2(4):951–972, 11 1992.

[2] OO Aalen. Effects of frailty in survival analysis. *Statistical Methods in Medical Research*, 3(3):227–243, 1994.

[3] Zeytu Gashaw Asfaw and Bo Henry Lindqvist. Unobserved heterogeneity in the power law nonhomogeneous poisson process. *Reliability Engineering & System Safety*, 134:59 – 65, 2015.

[4] Richard Barlow and Larry Hunter. Optimum preventive maintenance policies. *Operations Research*, 8(1):90–100, 1960.

[5] Laurence A. Baxter, Masaaki Kijima, and Michael Tortorella. A point process model for the reliability of a maintained system subject to general repair. *Communications in Statistics. Stochastic Models*, 12(1):12–1, 1996.

[6] F. Beichelt and K. Fischer. General failure model applied to preventive maintenance policies. *IEEE Transactions on Reliability*, R-29(1):39–41, 1980.

[7] Frank Beichelt. A unifying treatment of replacement policies with minimal repair. *Naval Research Logistics (NRL)*, 40(1):51–67, 1993.

[8] Henry Block, Yulin Li, and Thomas Savits. Initial and final behaviour of failure rate functions for mixtures and systems. *Journal of Applied Probability*, 40, 09 2003.

[9] Henry W. Block, Wagner S. Borges, and Thomas H. Savits. Age-dependent minimal repair. *Journal of Applied Probability*, 22(2):370–385, 1985.

[10] Henry W. Block, Yulin Li, Thomas H. Savits, and Jie Wang. Continuous mixtures with bathtub-shaped failure rates. *Journal of Applied Probability*, 45(1):260–270, 2008.

[11] S. Blumenthal, J. A. Greenwood, and L. Herbach. Superimposed non-stationary renewal processes. *Journal of Applied Probability*, 8(1):184–192, 1971.

[12] L. Bordes and Sophie Mercier. Extended geometric processes: Semiparametric estimation and application to reliability. *Journal of the Iranian Statistical Society*, 12:1–34, 03 2013.

[13] Willard Braun, Wei Li, and Y. Zhao. Properties of the geometric and related processes. *Naval Research Logistics (NRL)*, 52:607 – 616, 10 2005.

[14] Willard Braun, Wei Li, and Y. Zhao. Some theoretical properties of the geometric and α-series processes. *Communications in Statistics—Theory and Methods*, 37:1483–1496, 05 2008.

[15] Mark Brown and Frank Proschan. Imperfect repair. *Journal of Applied Probability*, 20(4):851–859, 1983.

[16] Cyril Caillault and Dominique Guegan. Empirical estimation of tail dependence using copulas. application to asian markets. *Quantitative Finance*, 5:489–501, 10 2005.

[17] Ji Cha and Maxim Finkelstein. Some notes on unobserved parameters (frailties) in reliability modeling. *Reliability Engineering & System Safety*, 123:99–103, 03 2014.

[18] Y. S. Chow and Herbert Robbins. A renewal theorem for random variables which are dependent or non-identically distributed. *Ann. Math. Statist.*, 34(2):390–395, 06 1963.

[19] David Clayton and Jack Cuzick. Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society. Series A (General)*, 148(2):82–117, 1985.

[20] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

[21] D. R. Cox and Walter L. Smith. On the superposition of renewal processes. *Biometrika*, 41(1/2):91–99, 1954.

[22] D.R. Cox and Valerie Isham. *The Theory of Stochastic Process.* Chapman and Hall, 1980.

[23] J. S. Dagpunar. Renewal-type equations for a general repair process. *Quality and Reliability Engineering International*, 13(4):235–245, 1997.

[24] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes.* Springer, 2003.

[25] Bram [de Jonge] and Philip A. Scarf. A review on maintenance optimization. *European Journal of Operational Research*, 2019.

[26] M. Deger, M. Helias, C. Boucsein, and S. Rotter. Statistical properties of superimposed stationary spike trains. *Journal of computational Neuroscience*, 32(3):443 – 463, 2012.

[27] Stefano Demarta and Alexander J. McNeil. The t copula and related copulas. *International Statistical Review / Revue Internationale de Statistique*, 73(1):111–129, 2005.

[28] Yann Dijoux, Mitra Fouladirad, and Dinh Tuan Nguyen. Statistical inference for imperfect maintenance models with missing data. *Reliability Engineering & System Safety*, 154:84 – 96, 2016.

[29] Theodosis Dimitrakos and E.G. Kyriakidis. An improved algorithm for the computation of the optimal repair/replacement policy under general repairs. *European Journal of Operational Research*, 182:775–782, 02 2007.

[30] L. Doyen. Reliability analysis and joint assessment of brown–proschan preventive maintenance efficiency and intrinsic wear-out. *Computational Statistics & Data Analysis*, 56(12):4433 – 4449, 2012.

[31] L. Doyen. Semi-parametric estimation of brown–proschan preventive maintenance effects and intrinsic wear-out. *Computational Statistics & Data Analysis*, 77:206 – 222, 2014.

[32] L. Doyen, R. Drouilhet, and L. Brenière. A generic framework for generalized virtual age models. *IEEE Transactions on Reliability*, pages 1–17, 2019.

[33] Laurent Doyen. Asymptotic properties of imperfect repair models and estimation of repair efficiency. *Naval Research Logistics (NRL)*, 57(3):296–307, 2010.

[34] Laurent Doyen and Olivier Gaudoin. Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering & System Safety*, 84(1):45 – 56, 2004. Selected papers from ESREL 2002.

[35] R. Drenick. The failure law of complex equipment. *Journal of the Society for Industrial and Applied Mathematics*, 8(4):680–690, 1960.

[36] Bradley Efron and David V. Hinkley. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika*, 65(3):457–482, 1978.

[37] M. Engelhardt and L.J. Bain. Statistical analysis of a compound power-law model for repairable systems. *IEEE Transactions on Reliability*, R-36(4):392–396, 1987. cited By 23.

[38] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley, 1968.

[39] Maxim Finkelstein. On some ageing properties of general repair processes. *Journal of Applied Probability - J APPL PROBAB*, 44, 03 2007.

[40] Maxim Finkelstein. *Failure rate modeling for reliability and risk*. Springer, 2008.

[41] Maxim Finkelstein. A note on converging geometric-type processes. *Journal of Applied Probability - J APPL PROBAB*, 47, 06 2010.

[42] Maxim Finkelstein. On the 'rate of aging' in heterogeneous populations. *Mathematical biosciences*, 232:20–3, 03 2011.

[43] Maxim Finkelstein. On the optimal degree of imperfect repair. *Reliability Engineering & System Safety*, 138:54 – 58, 2015.

[44] Maxim Finkelstein and Veronica Esaulova. Asymptotic behavior of a general class of mixture failure rates. *Advances in Applied Probability*, 38:244–262, 03 2006.

[45] Maxim Finkelstein and Mahmood Shafiee. Preventive maintenance for systems with repairable minor failures. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 231:101–108, 01 2017.

[46] M.S. Finkelstein. A scale model of general repair. *Microelectronics Reliability*, 33(1):41 – 44, 1993.

[47] Gabriel Frahm, Markus Junker, and Rafael Schmidt. Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, 37(1):80 – 100, 2005. Papers presented at the DeMoSTAFI Conference, Québec, 20-22 May 2004.

[48] Edward Furman, Alexey Kuznetsov, Jianxi Su, and Ričardas Zitikis. Tail dependence of the gaussian copula revisited. *Insurance: Mathematics and Economics*, 69:97 – 103, 2016.

[49] Christian Genest, Bruno Rémillard, and David Beaudoin. Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2):199 – 213, 2009.

[50] Gustavo L. Gilardoni, Maria Luiza Guerra [de Toledo], Marta A. Freitas, and Enrico A. Colosimo. Dynamics of an optimal maintenance policy for imperfect repair models. *European Journal of Operational Research*, 248(3):1104 – 1112, 2016.

[51] R. Guo and C. E. Love. Statistical analysis of an age model for imperfectly repaired systems. *Quality and Reliability Engineering International*, 8(2):133–146, 1992.

[52] Ramesh C. Gupta and Robin Warren. Determination of change points of non-monotonic failure rates. *Communications in Statistics - Theory and Methods*, 30(8-9):1903–1920, 2001.

[53] John Gurland and Jayaram Sethuraman. How pooling failure data may reverse increasing failure rates. *Journal of The American Statistical Association*, 90:1416–1423, 12 1995.

[54] J. Heckman and B. Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271–320, 1984.

[55] Myles Hollander, Brett Presnell, and Jayaram Sethuraman. Nonparametric methods for imperfect repair models. *Ann. Statist.*, 20(2):879–896, 06 1992.

[56] Rausand M Høyland A. *System Reliability Theory: Models, Statistical Methods, and Applications, 2nd Edition*. Wiley, 2003.

[57] J.H.Cha and M. Finkelstein. *Point Processes for Reliability Analysis*. Springer, 2018.

[58] R. Jiang. Two approximations of renewal function for any arbitrary lifetime distribution. *Annals of Operations Research*, pages 1089–1110, 8 2019.

[59] R. Jiang. A novel two-fold sectional approximation of renewal function and its applications. *Reliability Engineering & System Safety*, 193:106624, 2020.

[60] Xiaoyue Jiang, Viliam Makis, and Andrew Jardine. Optimal repair/replacement policy for a general repair model. *Advances in Applied Probability*, 33, 03 2001.

[61] Navarro Jorge and Pedro Hernandez. How to obtain bathtub-shaped failure rate models from normal mixtures. *Probability in the Engineering and Informational Sciences*, 18:511 – 531, 10 2004.

[62] M.J. Kallen. Modelling imperfect maintenance and the reliability of complex systems using superposed renewal processes. *Reliability Engineering and System Safety*, 96(6):636 – 641, 2011. ESREL 2009 Special Issue.

[63] E. L. Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[64] Samuel Karlin and Yosef Rinott. Classes of orderings of measures and related correlation inequalities ii. multivariate reverse rule distributions. *Journal of Multivariate Analysis*, 10(4):499 – 516, 1980.

[65] Hua Ke and Kai Yao. Block replacement policy with uncertain lifetimes. *Reliability Engineering & System Safety*, 148:119 – 124, 2016.

[66] Masaaki Kijima. Some results for repairable systems with general repair. *Journal of Applied Probability*, 26(1):89–102, 1989.

[67] Masaaki Kijima, Hidenori Morimura, and Yasusuke Suzuki. Periodical replacement problem without assuming minimal repair. *European Journal of Operational Research*, 37(2):194 – 203, 1988.

[68] Walter Kremers. An extension and implications of the inspection paradox. *Statistics & Probability Letters*, 6(4):269 – 273, 1988.

[69] Paul Kuehn. Approximate analysis of general queuing networks by decomposition. *Communications, IEEE Transactions on*, COM-27:113 – 126, 02 1979.

[70] Uday Kumar and Bengt Klefsjö. Reliability analysis of hydraulic systems of lhd machines using the power law process model. *Reliability Engineering & System Safety*, 35(3):217 – 224, 1992.

[71] C. Y. Teresa Lam and John P. Lehoczky. Superposition of renewal processes. *Advances in Applied Probability*, 23(1):64–85, 1991.

[72] Y. Lam. Geometric processes and replacement problem. 4:366–377, 1988.

[73] Y. Lam. *The geometric process and its applications*. World Scientific, 01 2007.

[74] Helge Langseth and Bo Henry Lindqvist. *A maintenance model for components exposed to several failure mechanisms and imperfect repair*, pages 415–430. World Scientific, 2004.

[75] Günter Last and Ryszard Szekli. Asymptotic and monotonicity properties of some repairable systems. *Advances in Applied Probability*, 30:1089–1110, 12 1998.

[76] Doyen Laurent. On the brown–proschan model when repair effects are unknown. *Applied Stochastic Models in Business and Industry*, 27:600 – 618, 11 2011.

[77] A. J. Lawrance. Dependency of intervals between events in superposition processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 35(2):306–315, 1973.

[78] E. L. Lehmann. Some concepts of dependence. *Ann. Math. Statist.*, 37(5):1137–1153, 10 1966.

[79] T.J. Lim. Estimating system reliability with fully masked data under brown-proschan imperfect repair model. *Reliability Engineering & System Safety*, 59(3):277 – 289, 1998.

[80] Bo Henry Lindqvist. *Statistical Modeling and Analysis of Repairable Systems*, pages 3–25. Birkhäuser Boston, Boston, MA, 1999.

[81] Xingheng Liu, Yann Dijoux, Jørn Vatn, and Håkon Toftaker. Performance of prognosis indicators for superimposed renewal processes. *Probability in the Engineering and Informational Sciences*, page 1–24, 2020.

[82] Xingheng Liu, Maxim Finkelstein, Jørn Vatn, and Yann Dijoux. Steady-state imperfect repair models. *European Journal of Operational Research*, 286(2):538 – 546, 2020.

[83] Xingheng Liu, Jørn Vatn, Yann Dijoux, and Håkon Toftaker. Unobserved heterogeneity in stable imperfect repair models. *Reliability Engineering & System Safety*, 203:107039, 2020.

[84] C.E. Love, Z.G. Zhang, M.A. Zitron, and R. Guo. A discrete semi-markov decision model to determine the optimal repair/replacement policy under general repairs. *European Journal of Operational Research*, 125(2):398 – 409, 2000.

[85] Viliam Makis and Andrew K.S. Jardine. A note on optimal replacement policy under general repair. *European Journal of Operational Research*, 69(1):75 – 82, 1993.

[86] Mazhar Ali Khan Malik. Reliable preventive maintenance scheduling. *A I I E Transactions*, 11(3):221–228, 1979.

[87] William Q Meeker and Luis A Escobar. *Statistical methods for reliability data*. New York : Wiley, 1998. "A Wiley-Interscience publication.".

[88] Chiranjit Mukhopadhyay and Mathews P. Samuel. Bayesian analysis of a superimposed renewal process. *Communications in Statistics - Theory and Methods*, 40(2):279–303, 2010.

[89] Dinh Tuan Nguyen, Yann Dijoux, and Mitra Fouladirad. Analytical properties of an imperfect repair model and application in preventive maintenance scheduling. *European Journal of Operational Research*, 256(2):439 – 453, 2017.

[90] Hossein Parsa and Mingzhou Jin. An improved approximation for the renewal function and its integral with an application in two-echelon inventory management. *International Journal of Production Economics*, 146(1):142 – 152, 2013.

[91] Hoang Pham and Hongzhou Wang. Imperfect maintenance. *European Journal of Operational Research*, 94(3):425 – 438, 1996.

[92] Konstadinos Politis and Susan M. Pitts. Approximations for solutions of renewal-type equations. *Stochastic Processes and their Applications*, 78(2):195 – 216, 1998.

[93] Frank Proschan. Theoretical explanation of observed decreasing failure rate. *Technometrics*, 5(3):375–383, 1963.

[94] Steven E. Rigdon and Asit P. Basu. The power law process: A model for the reliability of repairable systems. *Journal of Quality Technology*, 21(4):251–260, 1989.

[95] Sheldon M. Ross. *Stochastic Processes*. Wiley, 1995.

[96] Mahmood Shafiee and John Dalsgaard Sørensen. Maintenance optimization and inspection planning of wind energy assets: Models, methods and strategies. *Reliability Engineering & System Safety*, 192:105993, 2019. Complex Systems RAMS Optimization: Methods and Applications.

[97] M. Shaked and J.G. Shanthikumar. *Stochastic orders*. Springer Series in Statistics, 2007.

[98] Wujun Si, Ernie Love, and Qingyu Yang. Two-state optimal maintenance planning of repairable systems with covariate effects. *Computers & Operations Research*, 92:17 – 25, 2018.

[99] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8:229–231, 1959.

[100] Vaclav Slimacek and Bo Henry Lindqvist. Nonhomogeneous poisson process with nonparametric frailty. *Reliability Engineering & System Safety*, 149:14 – 23, 2016.

[101] Vaclav Slimacek and Bo Henry Lindqvist. Nonhomogeneous poisson process with nonparametric frailty and covariates. *Reliability Engineering & System Safety*, 167:75 – 83, 2017. Special Section: Applications of Probabilistic Graphical Models in Dependability, Diagnosis and Prognosis.

[102] W. L. Smith and M. R. Leadbetter. On the renewal function for the weibull distribution. *Technometrics*, 5(3):393–396, 1963.

[103] Wolfgang Stadje and Dror Zuckerman. Optimal maintenance strategies for repairable systems with general degree of repair. *Journal of Applied Probability*, 28(2):384–396, 1991.

[104] Chanan S. Syan and Geeta Ramsoobag. Maintenance applications of multi-criteria optimization: A review. *Reliability Engineering & System Safety*, 190:106520, 2019.

[105] P. Torab and E. W. Kamen. On approximate renewal models for the superposition of renewal processes. In *ICC 2001. IEEE International Conference on Communications. Conference Record (Cat. No.01CH37240)*, volume 9, pages 2901–2906 vol.9, 2001.

[106] Catalina A. Vallejos and Mark F.J. Steel. Incorporating unobserved heterogeneity in weibull survival models: A bayesian approach. *Econometrics and Statistics*, 3:73 – 88, 2017.

[107] J. Vatn. Heterogeneity of weibull samples. In *Probabilistic Safety Assessment and Management 96*, 1996.

[108] James Vaupel and Anatoliy Yashin. Repeated resuscitation: How lifesaving alters life tables. *Demography*, 24:123–35, 03 1987.

[109] James W. Vaupel, Kenneth G. Manton, and Eric Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454, Aug 1979.

[110] Hongzhou Wang and Hoang Pham. *Reliability and Optimal Maintenance*. Springer, 01 2006.

[111] Lyn R. Whitaker and Francisco J. Samaniego. Estimating the reliability of systems subject to imperfect repair. *Journal of the American Statistical Association*, 84(405):301–309, 1989.

[112] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25, 1982.

[113] Ward Whitt. Approximating a point process by a renewal process, i: Two basic methods. *Operations Research*, 30(1):125–147, 1982.

[114] Andreas Wienke. *Frailty Models in Survival Analysis*. CRC Press, 2007.

[115] James R. Wilson. The inspection paradox in renewal-reward processes. *Operations Research Letters*, 2(1):27 – 30, 1983.

[116] Shaomin Wu and Philip Scarf. Two new stochastic models of the failure process of a series system. *European Journal of Operational Research*, 257(3):763 – 772, 2017.

[117] Olexandr Yevkin. A monte carlo approach for evaluation of availability and failure intensity under g-renewal process model. In *Proceedings of ESREL conference*, pages 1015–1021, 08 2011.

[118] Wei Zhang, Ye Tian, Luis A. Escobar, and William Q. Meeker. Estimating a parametric component lifetime distribution from a collection of superimposed renewal processes. *Technometrics*, 59(2):202–214, 2017.

Xingheng LIU

Doctorat : Optimisation et Sûreté des Systèmes

Année 2020

Modèles de maintenance imparfaite : état stationnaire, hétérogénéité et applications

Au fil des ans, la maintenance a attiré l'attention dans de nombreux domaines. Les défaillances des systèmes peuvent entraîner des pertes de la production ou des accidents majeurs. Par conséquent, au lieu de se concentrer sur la maintenance réactive, qui vise à restaurer un système après une panne inattendue, les gens sont plus intéressés par les maintenances préventives et prédictives qui pourraient nous préparer à la panne ou éviter un scénario catastrophique. Pour faire un plan de maintenance rentable, il faut décider quelles sont les actions de maintenance appropriées et quand les mettre en œuvre. Une compréhension approfondie du mécanisme de défaillance du système est donc nécessaire. De nombreux modèles mathématiques ont été proposés pour illustrer le processus de défaillance/réparation qui subit des réparations imparfaites. Dans ce contexte, nous étudions le domaine d'application des modèles de réparation imparfaits tout en examinant les conséquences d'un ajustement de modèle inapproprié, et proposons de nouveaux modèles que nous pensons plus réalistes. Les propriétés mathématiques sont étudiées et l'inférence statistique est abordée. Les développements théoriques sont accompagnés de plusieurs études de cas, basées sur des données collectées du réseau ferroviaire norvégien. Nous avons l'intention de souligner la valeur des modèles de réparation imparfaits pour évaluer la fiabilité des systèmes, évaluer l'efficacité des réparations et optimiser les plans de maintenance. Nous espérons que ce travail sera utile et inspirant pour les praticiens et chercheurs concernés.

Mots clés : entretien – fiabilité – durée de vie (ingénierie) – processus stochastiques – estimation de paramètres.

Imperfect Maintenance Models: Steady-state, Heterogeneity and Applications

Over the years, maintenance has been gaining attention in many different fields. Failures of systems can result in as small as the loss of production or as large as a major accident. Consequently, instead of focusing on reactive maintenance, which aims to restore a system after an unexpected failure, people are more interested in preventive and predictive maintenances that could prepare us for the failure or avoid a catastrophic scenario. To make a cost-effective maintenance plan, one must decide what the appropriate maintenance actions are, and when to implement them? A profound understanding of the system's failure mechanism is necessary to answer these questions. Numerous mathematical models have been proposed to depict the failure/repair process that undergoes imperfect repairs. The thesis is completed in this context. On the one hand, we investigate the application field of imperfect repair models while examining the consequences of inappropriate model fitting. On the other hand, new models, which we believe are more realistic, are established. The mathematical properties are thoroughly studied, and statistical inference is addressed. The theoretical developments are accompanied by several case studies, based on not only simulated data but also data collected from the Norwegian railway network. We intend to highlight the value of imperfect repair models in evaluating the reliability of systems, assessing the effectiveness of repairs, and optimizing maintenance plans. We hope that this work would be helpful and inspiring for relevant practitioners and researchers.

Keywords: maintenance – reliability – service life (engineering) – stochastic processes – parameter estimation.

Thèse réalisée en partenariat entre :

Union européenne
Fonds Européen de Développement Régional

l'Europe s'engage en Champagne-Ardenne avec le FEDER

NTNU
Norwegian University of Science and Technology

utt
Université de Technologie Troyes

Ecole Doctorale "Sciences pour l'Ingénieur"