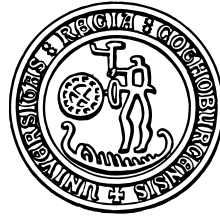


THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Survival Estimation and Distribution Approximation for Optimal Maintenance

JOHAN SVENSSON

CHALMERS | GÖTEBORG UNIVERSITY



Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg, Sweden
Göteborg, Sweden 2004

Survival Estimation and Distribution Approximation for Optimal Maintenance

JOHAN SVENSSON

©JOHAN SVENSSON, 2004

ISSN 0347-2809/No 2004:33

Department of Mathematical Statistics

Chalmers University of Technology and Göteborg University

SE-412 96 Göteborg, Sweden

Telephone +46 (0)31-772 1000

Göteborg, 2004

Survival Estimation and Distribution Approximation for Optimal Maintenance

JOHAN SVENSSON

Department of Mathematical Statistics

Chalmers University of Technology and Göteborg University

Abstract

The problem of rational maintenance of aircraft engines is studied with respect to the influence of random events. The starting point is an optimization model suggesting what parts in the engine that should be replaced at each maintenance time. The input data is the state of the details in the engine. Two models are developed that estimate life distributions of these details. The first model is a non-stationary renewal process and the second model is a non-homogeneous Poisson process. Real data consisting of times between repairs is used and a measure to compare the models is defined. With our data the non-stationary renewal process works better. Different repair stations affect the life of the components, which the non-stationary renewal process manages to model. This model also manages the aging component problem in an effective way. However, in this case no aging is present other than substantial degeneration after the first repair.

In order to get an interface with the optimization model the distributions need to be discrete. Four methods to make discretizations are discussed and adapted to suit the model. The methods are compared and the choice concerning the number of points of support is discussed. Finally the consequence of using a narrow scenario tree is commented upon.

Keywords: non-stationary renewal process; non-homogeneous Poisson process; survival; optimal maintenance; discretization; points of support

Acknowledgements

The author wants to thank Dragi Anevski at Göteborg University and Thomas Svensson at Fraunhofer Chalmers Center and Jacques de Maré at Chalmers for supervision. Furthermore I own a debt of gratitude to Fredrik Plym and Maud Österman at Volvo Aero Corporation. This work was financed by NFFP.

Contents

List of papers	vi
1 Introduction	1
2 The maintenance process	2
3 Existing Data	3
3.1 Flameholder	4
3.2 Improvements of future data	5
4 Survival analysis	5
5 Summary of Paper A	7
6 Scenarios and stages in stochastic optimization	8
7 Summary of Paper B	10
8 Future work	11

List of papers

This thesis includes the following papers

Paper A J.Svensson *Two statistical models used on aircraft engine data modeling times between repairs.*

Paper B J.Svensson *Discrete approximations of life distributions in optimal replacement.*

1 Introduction

Aircraft engines can be more economically maintained and resources can be saved if the maintenance process is improved. In order to obtain a better maintenance process, an optimization which indicates what parts in the engine should be replaced is being developed. The criterion for optimality is the expected cost to maintain the engine. Input data for the optimization model contains a description of the state of the deterministic details and the stochastic details.

Definition 1: A *deterministic detail* is a component that has a predetermined limited time in service which must not be exceeded.

Deterministic details are components that are vital for safety. If a deterministic component fails there is a risk that the engine will stop functioning. In practice the upper limit on the service time means that most deterministic parts will be replaced long before they are even close to failure.

Definition 2: A *stochastic detail* is a component that is not deterministic.

Stochastic components are allowed to operate in the aircraft engine until they break. If a stochastic component breaks during a flight the engine will still work but with reduced performance.

In the Volvo military engines RM8 and RM12 there are about 50 components that are not expected to last the engines pre-specified life, among these about 40 are deterministic and 10 are stochastic. More than 40% of all cases when the engines are taken to be repaired are unplanned, triggered by the failure of a stochastic component.

A model that describes the properties of the stochastic components is needed. Such a model is developed with the type of data that exists at Volvo Aero Corporation (VAC). To get the stochastic components to interface with the optimization model, approximations need to be made. How these things are connected is shown in Figure 1.

In the first paper, Paper A, two models are presented that predict the time to next failure for stochastic components. The first model is a Non-Stationary Renewal Process and the second model is a Non-Homogeneous Poisson Process. The models are illustrated on a set of data from the engine RM8. An analysis of how different repair stations affect the life span of the components is done with the help of the Non-Stationary Renewal Process. Finally a way of modeling aging components is suggested.

The second paper, Paper B, is about the interface between the stochastic components and the optimization model. In order to compare different

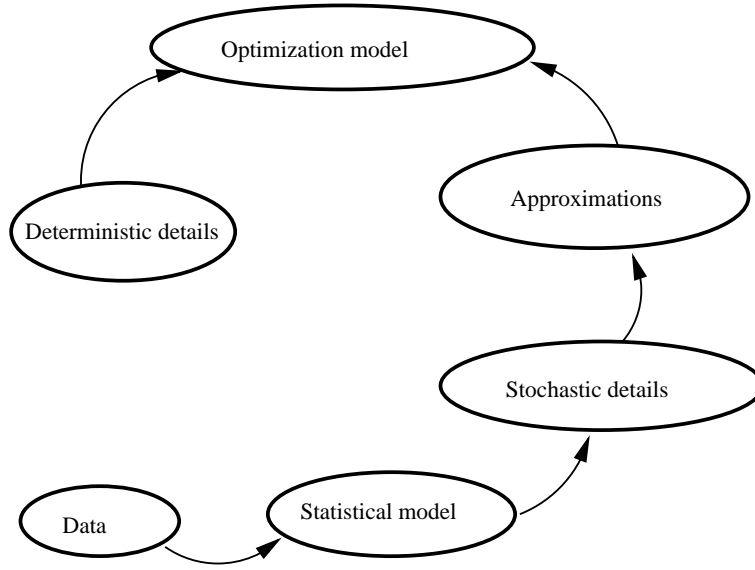


Figure 1: *Illustration of how the models in this thesis are connected.*

approximations an error measure is introduced. Different methods of doing discretizations are discussed and adapted to be used with the constraints that the optimization model requires. Finally the consequence of using a narrow scenario tree is commented upon.

The outline of this thesis is as follows. In Sections 2 and 3 some background information about the maintenance process at VAC is supplied. The nature of the data that exists is given, and some parameters on the flameholder, which is a detail at the far end of the engine, is presented. Some suggestions of possible improvements of future data is mentioned. Sections 4 and 5 are summaries of papers A and B. In Section 6 we discuss future work.

2 The maintenance process

There are two philosophies at VAC when it comes to maintenance: either maintenance is performed at certain fixed times or maintenance is guided by (known) requirements.

If the engine is obtained at certain predetermined times the maintenance times are known in advance but not the parts which are to be replaced.

If the maintenance process is ruled by the second philosophy, the engine is obtained for control only if there are any signs of failure or lacking performance. In this case neither maintenance times nor maintenance needs are known in advance. When the aircraft lands and the engine is still near the hangar it is examined for broken components. A component is considered broken if it fails to comply with a set of fitness rules when it is observed. If a component is broken the entire engine or possibly a module of the engine is sent to a repair station. Not all components are examined after each flight so in practice this means that the components are checked at certain intervals. These intervals are typically shorter than the predetermined intervals used in the first philosophy.

This implies that we do not know the exact time of the failure. It can be anytime between two observations. Some components are observed every time the aircraft lands. Since the aircraft mission is short in comparison to the life of the component, the time when those components are observed and determined as broken can be considered as the time of failure.

The older RM8 military engine was mostly maintained by the first philosophy but some components were checked after every flight mission. The plan is that the newer RM12 military engine should be maintained according to the second philosophy.

3 Existing Data

The data set available for our analysis consists of historical data on the engines RM8 and RM12. The engines have been carefully followed up. Information about every component in each engine is gathered in a database in which each component has a unique individual number. Relevant information that exists about failures is how, where and when the failures are detected. There is information about what engine the failing component was operating in and what countermeasures were implemented, repair or replacement. There is also a description of why the component was considered broken. Components that are currently in service and hence can be considered as censored observations are also possible to keep track of.

To facilitate further analyses, it was suitable to divide the data on RM8 into two sets RM8a and RM8b. The RM12 and RM8b are registered in the system RUF, RM8a is not. In the RUF system, aircraft data is sampled during operation, with a frequency of 7.5 Hz. After the flight the information is downloaded to a ground station, where some data reduction is done and the information stored. Examples of the type of information sampled during flight are turbine rotation speeds, temperatures and pressures; common parameters stored at the ground station are flight time and a measure of equivalent low cycle fatigue (ELCF).

The older engine RM8 has been used for a long time and a great amount of data exists. The newer RM12 engine has experienced fewer flight hours and the dataset of failures is very small. Most stochastic components have not experienced any failure yet.

On information of the kind of existing data, a decision was made to create a model of the RM8 engine and then try to transfer this model to RM12.

3.1 Flameholder

In order to test the models the flameholder component was chosen. The flameholder is a stochastic component at the far end of the engine in the afterburner module. In the RM8 the flameholder consists of two rings. The RM12 flameholder consists of only one part. The task of the flameholder is to hold the flame and put the passing air in rotation and create recirculation zones with low axial circulation speed, important for the ignition of the fuel.

The most common failures of the afterburner are burnouts, when parts of the flameholder have been burnt away, and cracks. The flameholder is checked after every flight mission.

Parameters in the RUF system that we believe have some impact on the life on the flameholder are:

- Time with afterburner ignition (TAI).
- Time with max afterburner ignition (TMAI).
- Number of afterburner ignitions (NAI).

Other parameters that are always available are:

- Total operation time (TOT), the pilots judgment of the flight time registered in a log.
- Engine time (ET), the time the engine is rotating with a speed greater than 140 radians per second.
- Flight time (FT), triggered by the landing gear.

Correlations between these variables for the RM12 engine are given in the correlation matrix

$$\Sigma = \begin{bmatrix} (TOT) & (TAI) & (TMAI) & (NAI) & (TE) & (FH) \\ 1.00 & 0.94 & 0.92 & 0.96 & 0.98 & 0.99 \\ 0.94 & 1.00 & 0.99 & 0.97 & 0.94 & 0.94 \\ 0.92 & 0.99 & 1.00 & 0.96 & 0.93 & 0.92 \\ 0.96 & 0.97 & 0.96 & 1.00 & 0.95 & 0.96 \\ 0.98 & 0.94 & 0.93 & 0.95 & 1.00 & 0.98 \\ 0.99 & 0.94 & 0.92 & 0.96 & 0.98 & 1.00 \end{bmatrix}.$$

As we can see the parameters (TOT), (TE) and (FH) have a very high correlation, not surprisingly since they are all different measures of the time the engine is operating. The other parameters (TAI), (TMAI) and (NAI) are highly correlated, partly among themselves but also to the times (TOT), (TE) and (FH). We will therefore limit ourselves to studying only one parameter.

3.2 Improvements of future data

In order to make better predictions and more accurate models, improvements in the data gathering and storing can be made.

When an engine for one reason or other is entering a repair station it is disassembled. In the disassembling process more parts that do not comply with the set of fitness rules may be discovered. Therefore there is a connection between the parts that forced the engine to the repair bay and other parts that are replaced. If information about this could be gathered components with high connection could maybe be consolidated in the optimization model.

Some components can experience different types of failures. If data on these was stored, it would be possible to detect if different failures have different failure distributions. Since there are fitness rules, a failure can break one or more rules. A code describing the error type would be a good feature. Currently the description is written in text.

Data consists of accumulated numbers of events that a detail has experienced from creation to death or censoring, no matter if the detail has been repaired several times. Better analyses could be performed if we knew the accumulated value of the parameter at every repair.

Storing more than the accumulated value of the parameters would facilitate the use of more advanced models, cf. e.g. Roemer and Ghiocel (11) and Roemer and Kacprzyński (12). The advantage of the extension to those ideas is that the information collected during flight can directly predict the remaining life. Possible problems with such models is the high level of noise and variation in data, cf. Krok and Ashby (8).

4 Survival analysis

The main goal of survival analysis is to determine the survival time of a population of individuals with the help of data. The individuals can be people, animals, components, etc. The methods for estimating survival can be divided into parametric and non-parametric methods. Parametric methods are methods that assume that the survival follows a certain dis-

tribution described by a number of parameters. Non-parametric methods assume less about the distribution.

A way of finding a parametric distribution that fits the data well is to first use a non-parametric estimate of the distribution and see what it looks like. Then a parametric distribution that is close to the non-parametric estimate can be chosen. Standard non-parametric estimators are the Nelson Aalen estimator and the Product-Limit estimator, also called Kaplan-Meier estimator. The Nelson-Aalen estimator estimates the cumulative hazard rate while the Product-Limit estimator estimates the survival function $S(t) = 1 - F(t)$ where $F(t)$ is the cumulative distribution function.

The cumulative hazard rate is defined as

$$H(t) = \int_0^t h(u) du,$$

where

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t},$$

where T is the stochastic variable we wish to describe. Nelson-Aalen estimator is defined as

$$\hat{H}(t) = \begin{cases} 0 & \text{if } t < t_1 \\ \sum_{t_i \leq t} \frac{1}{Y(t_i)} & \text{if } t \geq t_1 \end{cases}$$

and the Product-Limit estimator is defined as

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{i: t_i \leq t} (1 - \frac{d_i}{Y(t_i)}) & \text{if } t \geq t_1, \end{cases}$$

where t_i are the ordered observation times and $Y(t_i)$ are the number of individuals at risk at time t_i and d_i the number of failures at time t_i .

There exist several parametric distributions that are known to be well suited for describing survival. Some of the most common are the Weibull distribution

$$F(t) = 1 - e^{-(\frac{1}{\theta}t)^\alpha}, \quad t > 0, \quad (\theta > 0, \alpha > 0),$$

and the Log-normal distribution

$$F(t) = \Phi\left(\frac{\log t - \mu}{\sigma}\right), \quad t > 0, \quad (\sigma > 0).$$

The Maximum Likelihood estimator is a common parametric estimator that can handle different kinds of truncation and interval censored observations. In our application the Maximum Likelihood function L can be written

$$L(\theta) = \prod_{i \in D} f_\theta(t_i) \prod_{j \in R} S_\theta(c_j) \prod_{k \in I} (S_\theta(l_k) - S_\theta(r_k))$$

where D is the set of all failure times, R is the set of right-censored observations, and I is the set of all interval-censored observations. If individual k is interval-censored we know that individual k has broken between times l_k and r_k . The parameter vector θ is estimated as

$$\theta = \underset{\theta}{\operatorname{argmax}} L(\theta).$$

A comparison between the parametric and non-parametric estimates can be made by comparing their hazard rates through hypothesis testing, see Moeschberger and Klein (7). A drawback with such a comparison is that the parametric distribution is often rejected if the data set is large. The true distribution of the individuals follows no common parametric distribution but has its own distribution. This does not necessarily mean that the parametric distribution is a bad model. This must be taken into account when deciding if the parametric distribution is a good enough model for our application. A visual comparison can be made in several ways. One way is to compare the density functions between the non-parametric and parametric estimation. The non-parametric estimation of the density function can be obtained by performing a kernel smoothing, see further Moeschberger and Klein (7).

5 Summary of Paper A

Military aircraft engines can offer greater operational availability and be more economically maintained through the use of better models to predict times to failure. Two models are used to analyze data gathered from Volvo Aero Corporation in Trollhättan. We are interested in the failure time distribution of the flameholder in the new RM12 engine. We have limited knowledge about it due to the limited number of RM12 engines currently in service. We have a large data set containing repair and maintenance times for the same type of details in the older RM8 engine. This paper will not discuss how to transform knowledge of the RM8 engine to the RM12 engine but will instead predict repair or maintenance times for the RM8 engine.

The first model is a non-stationary renewal process (NSRP) and the second is a non-homogeneous Poisson process (NHPP). We are interested in estimating the survival function and the hazard rate. In the NSRP we make a non-parametric estimation of the survival function with the help of the Product-Limit estimator, cf. Høyland and Rausand (5). We use kernel smoothing, cf. Klein and Moeschberger (7), to make a visual illustration of the density function. When the density function is estimated a parametric model is chosen to describe the times between repairs. Different estimation

methods are discussed. In the NHPP we use the Nelson Aalen estimator to estimate the cumulative hazard function, cf. Andersen et.al. (1). The NHPP is using the minimal repair assumption, cf. Høyland and Rausand (5), while in the NSRP the time to first repair is independent of the time to second repair.

An error measure is defined to compare how well the different models are suited to model current data, and we conclude that the NSRP process is the better model.

The NSRP process is developed to model different kinds of repair stations. There is one repair station at VAC and smaller repair stations closer to the hangars. If engines were repaired at the repair station at it took longer time for VAC turned subsequent failures to occur. Finally two tests were made to see if the components are aging, but no aging was present other than a substantial degeneration after the first repair.

6 Scenarios and stages in stochastic optimization

Stochastic optimization occurs when one or more variables in the optimization model are stochastic. Most stochastic problems can be formulated

$$z = \min_{x \in X} \int f(x, u) dG(u) \quad (1)$$

where U is a stochastic variable and G is the distribution of U and X some restriction of x . Problems of form (1) are often too hard to solve and simplifications are needed. One common approach is to create scenarios u_i which represent some outcomes of the stochastic variable U . To each outcome belongs a probability p_i . Since the scenarios can not cover all the outcomes of a continuous stochastic variable, this is an approximating method. However the formulation (1) is changed to

$$Z = \min_{x \in X} \sum f(x, u_i) p_i$$

which is easier to solve, especially if f is a linear function. The scenarios are structured in a scenario tree. An example of a scenario tree with a two dimensional stochastic variable $U = (R_1, R_2)$ is shown in Figure 2. The random variables R_1 and R_2 can e.g. describe the lives of two stochastic components in an aircraft engine. Here five scenarios that represent possible outcomes for U are chosen. The outcomes chosen in the tree must be representative of the entire spectrum of outcomes.

Modeling over time where decisions have to be made on several occasions makes it natural to introduce stages. Assume we have to make decisions

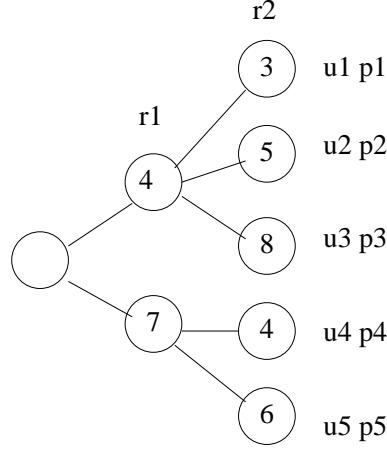


Figure 2: A scenario tree with three scenarios for R_2 if $R_1=4$ and two if $R_1 = 7$. A total of five scenarios u_1, \dots, u_5 with probabilities p_1, \dots, p_5 .

today that influence the decisions tomorrow that affect the future. We can then introduce a multistage model with three stages: today = 0, tomorrow = 1 and future = 2. Let f_0 denote our cost function today and let $f_{1|x_0}$ denote the cost function tomorrow which depends on x_0 , the decision today. The cost function in the future $f_{2|x_0, x_1}$ depends on both the decisions we make today and tomorrow. In this case the three stage problem can be written.

$$z = \min_{x_0 \in X_0} f_0(x_0) \quad (2)$$

where

$$f_0(x_0) = \min_{x_1 \in X_1(x_0)} f_{1|x_0}(x_1) \quad (3)$$

and

$$f_{1|x_0}(x_1) = \min_{x_2 \in X_2(x_0, x_1)} f_{2|x_0, x_1}(x_2) \quad (4)$$

where x_1, x_2, x_3 are variables and X_1, X_2, X_3 are the domains. If we have no stochastic variables or if we make a scenario tree (discretization of continuous random variable), equations (2), (3) and (4), can be rewritten as

$$z = \min_{\mathbf{x} \in \mathbf{X}} \tilde{f}(\mathbf{x}) \quad (5)$$

where $\mathbf{x} = (x_0, x_1, x_2)$ are the variables and \mathbf{X} is the allowed domain for \mathbf{x} . Using the forms (2), (3) and (4) often makes the problem easier to model

and understand. Sometimes it is harder to solve (5) than the connected problems (2), (3) and (4).

The optimization model used in this thesis has integer restrictions on the x variables. A problem with integer restrictions is much harder to solve than the problem arising if the integer restrictions are removed. For further discussion of stochastic optimization with scenarios and stages, see Kall and Wallace (6).

7 Summary of Paper B

Aircraft engines can be more economically maintained and resources can be saved if the maintenance process is optimal. It can be a hard decision to decide what components in an engine to replace when the engine is being maintained. Several optimization models have been developed to deal with this problem. Epstein and Wilamowsky (3) and Dickman et al. (4) have developed models for modeling components with predetermined deterministic lives. Andréasson (2) has developed a model for details with deterministic lives as well as details with stochastic lives. Another approach to the problem is to construct a maintenance policy that is not always optimal but hopefully good. A survey of replacement and maintenance policies can be found in Wang (14).

In this paper we will use the model presented in Andréasson (2) and formulate it as a two-step model. In order to use stochastic components in the optimization, the density functions need to be in a discrete form and there are also restrictions on what points of supports that are allowed. An error measure closely related to the model is formulated. The error measure measures the cost of using different kinds of discretizations and the size of the error is related to the sup-distance between the distribution and discretization. Four discretization methods will be presented and adapted to the constraints. The first method is the bracket mean method, cf Smith (13), the second method minimizes the Wasserstein distance, cf. Pflug (10) and the third method keeps the moment of the distribution, cf. Miller and Rice (9). The last method is a method that minimizes the sup-distance.

In order to keep down calculation times we want as few points of support as possible. Test runs with a Weibull distribution are done to compare the different discretization methods and conclude how the choice of number of points of support affects the accuracy. In the test the method that minimizes the Wasserstein and Sup-distance performed best. The method that preserved the moments performed worse. The error decreases with the number of support points. The minimum number of points of support that is suggested is three.

Finally the consequence of using a narrow scenario tree is discussed. A

component that is exponentially distributed is sometimes replaced although the component is not aging.

8 Future work

In order to get the optimization model to work with the stochastic components from RM12 more work needs to be done. The following are some suggestions for future research.

An investigation of how to translate the distributional properties of the life of the RM8 engine to the RM12 engine by e.g. a Bayesian approach.

The lives of the stochastic components may be dependent among themselves but also affected by the age of deterministic components and the age of the engine itself. It is not obvious how to get this information from the existing data. There is also dependence between parts because they were observed when the engines were sent in for repairing other failures.

In the current approach only the first failure of a stochastic component in the optimization model is modeled with several points of support, the remaining times to failure is modeled as deterministic and equal to the expected value of the time to failure. Relaxing this restriction would give more reliable suggestions of what parts to replace but also increase the calculation time.

The relevant details in an engine can be divided into 40 deterministic parts and 10 stochastic parts. More work needs to be done when it comes to formulating an optimization model and finding approximations that allow the optimization problem to be solved fast, when we have more than one stochastic component.

References

- [1] P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, 1993.
- [2] N. Andréasson. Optimization of opportunistic replacement activities in deterministic and stochastic multi-component systems. Licentiate thesis, Chalmers University of Technology, 2004.
- [3] Y. Epstein, S. Wilamowsky. Opportunistic replacement in a deterministic environment. *Comput. and Opns. Res.*, 12:311–322, 1985.
- [4] B. Dickman et al. Modeling deterministic opportunistic replacement as an integer programming problem. *Amer. J.Math. and Mgmt.*, 11, 1991.

- [5] A. Høyland and R. Rausand. *System Reliability Theory*. John Wiley & Sons Inc., 1994.
- [6] Peter Kall and Stein W. Wallace. *Stochastic programming*. Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons Ltd., Chichester, 1994.
- [7] J. Klein and M. Moeschberger. *Survival Analysis*. Springer-Verlag Inc., New York, 1997.
- [8] M. J. Krok and M. J. Ashby. Condition-based diagnostic gas path reasoning for gas turbine engines. In *Proceeding of the 2002 IEEE International Conference on Control Applications*, pages 227–234. 2002.
- [9] T.R. Miller, A.C. Rice. Discrete approximations to probability distributions. *Management Science*, 29:352–362, 1983.
- [10] G. Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Math. Program.*, 89(2, Ser. B):251–271, 2001. Mathematical programming and finance.
- [11] M.J. Roemer and D.M. Ghiocel. Probabilistic approach to the diagnosis of gas turbine engines. Technical report, 53rd Machinery Prevention Technologies (MFPT), Conference, Virginia Beach, VA, April 1999.
- [12] M.J. Roemer and G.J. Kacprzyński. Advanced diagnostics and prognostics for gas turbine engine risk assessment. Technical report, IGTI/ASME Turbo Expo, Munich, Germany, May 2000.
- [13] J.E. Smith. Moment methods for decision analysis. *Management Science*, 39:340–358, 1993.
- [14] H. Wang. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139:469–489, 2002.

Paper A

Two statistical models used on aircraft engine data modeling times between repairs

Johan Svensson
Chalmers University of Technology
Department of Mathematical Statistics
SE-41296 Göteborg Sweden

Abstract

Military aircraft engines can offer greater operational availability and be more economically maintained through the use of better models that predict times to failure. In this paper, real data consisting of times between repairs of a flame holder in an aircraft engine is used and two models that predict the time to next failure are suggested. The first model is a non-stationary renewal process and the second model is a non-homogeneous Poisson process. A measure to compare the models is defined and with our data the non-stationary renewal process works better. Different repair stations affect the life span of the components but the non-stationary renewal process manages to model this. This model also manages the aging component problem in a effective way. However, in this case no aging is present other than substantial degeneration after the first repair.

Keywords: non-stationary renewal process; non-homogeneous Poisson process; survival;

1 Introduction

Military aircraft engines can offer greater operational availability and be more economically maintained through the use of better models that predict times to failure. Some models exist that use different strategies. Roemer and Ghiocel (12) describe an interesting model that is based on a resistance variable R and a stress variable S . When $R - S < 0$ failure occurs. Tinga and Visser (15) suggest a model that involves a fatigue model, a gas turbine simulation program, fluid dynamic model and a finite element model. This is combined with a statistical reliability model. Roemer and Ghiocel (13) present a set of tools for health monitoring, diagnostic and prognostic

of turbo machinery. One nice tool define two indices and use multivariate process control methods to monitor engines. They also suggest a simple fatigue model for individual components. All of the above models require continuous monitoring of several parameters that are relevant for the life span of the component. They also require that the parameters have historically been stored in such a way that it is possible to make any use of them. Krok and Ashby (10) states that it is hard to develop models based on most of the monitored parameters due high levels of noise and large variation in data.

In other parts of the literature, hundreds of repair and maintenance models can be found. Basic mathematical models and methods can be found in, for instance Høyland and Rausand (5), Klein and Moeschberger (9), Kalbfleisch and Prentice (6) or Andersen et al. (1). A statistical model that involves physical laws can be found in Yang (19). Also, pure statistical models can be found in Kaminskiy (7) or Kijima and Masaaki (8). A survey of some maintenance and reliability models can be found in Weiss, George H. (18).

Based on existing models, this paper will present two models in the class of non-stationary renewal process and non-homogeneous Poisson process to predict failures. The models use historical data for calibration and are suitable for modeling time to failure of stochastic components. A stochastic component is a component that is allowed to operate in the aircraft engine until it breaks. If the component breaks under a flight mission the engine will still work but with reduced performance. A deterministic component is a component that is only allowed to function in the aircraft engine a predetermine specific time. If a deterministic component breaks under a flight mission there is a risk that the engine will stop functioning. In practice this means that the deterministic component is replaced long before its life span is consumed. Methods for calculating the preset times for deterministic components are not discussed here.

There are only two states a component can be in, either functioning or broke. Hence we have a two state system. Recently a lot of work have been made on multistage systems, a survey of multistage system can be found in El-Newehi and Proschan (3). A detail is considered broken if it fails to comply with a set of fitness rules when it is observed. The flight mission time is short and the details studied here are observed after every flight mission and therefore the time to failure is considered to be the time when the component is observed and fails to comply with the set of fitness rules. In other scenarios, mission times may not be short, or the details are just observed on a few special occasions due to the fact that they are positioned deep inside the engine. The method discussed can still be applied but the derivation of the estimations of the distribution functions may be slightly

different.

The data used in this work is gathered from Volvo Aero Corporation in Trollhättan and we are interested in the failure distribution of the flame holder in the new RM12 engine. We have limited knowledge about this due to the limited number RM12 engines currently in service. We have a large data set containing repair and maintenance times for the same type of details in an older version of the RM8 engine. This data will be of some help in predicting maintenance times for the RM12 engine. This paper will not discuss how to transform knowledge of the RM8 engine to the RM12 engine but will instead discuss two models that predict repair or maintenance times for the RM8 engine.

In the following sections, we present the data material and define two different models. We discuss existing methods to analyze the unknown parameters in the models. An error measure is defined and a comparison between the models is made, using this measure. We modify one of the models so that it can handle different kinds of repairs. Data from the older version of the engine is used. We believe that the models when applied to the new engine design will help us understand the life of the newly designed details.

2 The models

In the literature there are many models, both simple and more complex, that can be applied to predict the time to the next repair. Here we use two models that are in the classes of NSRP (*Non-Stationary Renewal Processes*) respectively NHPP (*Non-Homogeneous Poisson Processes*) to analyse data. We discuss existing methods to analyse the unknown parameters in the models, and close the section with a discussion on measures of goodness of fit.

2.1 The set of data

The data used in this paper originate from a detail called the flame holder which is an essential part in the after burner system. The flame holder consists of two parts called the inner ring and outer ring. The outer ring exists in two versions. We call the inner ring detail 1 and the outer ring detail 2 and 3. We observe the times between repairs. Every time a component fails, a decision is made whether the component should be repaired or scrapped. For every component i we observe a sequence of times between repairs $\{T_1^i, \dots, T_{k_i}^i\}$, $i = 1, \dots, n$, where the last observation possibly is censored, meaning that for the last observation we may have the information

that the time to next repair is longer than the time observed. For more information on censoring see Klein and Moeschberger (9). We model different components as independent, that is $\{T^i\}$ and $\{T^j\}$ are independent vectors if $i \neq j$. We also assume that the censoring process is noninformative, see i.e Kalbfleisch and Prentice (6).

We have more observations of repairs in details 2 and 3 than in detail 1. Doing simple descriptive statistics, we get histogram plots of the number of repairs for each component in Figure 1. The x-axis corresponds to the number of repairs and the y-axis to the number of components that have been repaired x times.

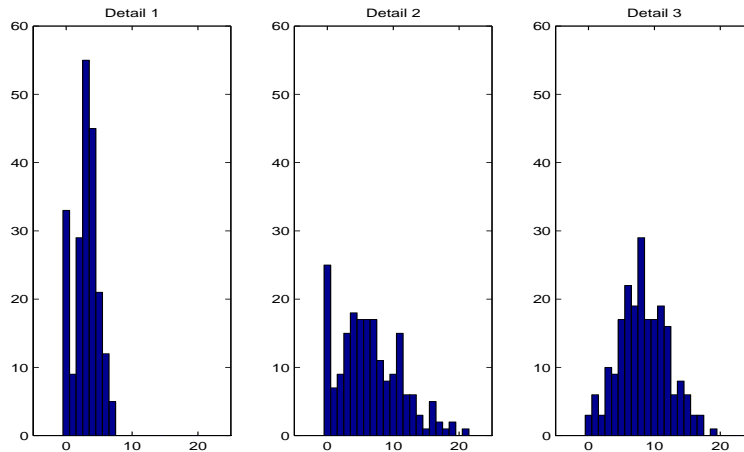


Figure 1: *Histogram on repaired components of each detail. The x-axis corresponds to the number of repairs and the y-axis number of components that have been repaired x times.*

2.2 NSRP-model

A non-stationary renewal process (NSRP) is defined in the following way (cf. e.g. Høyland and Rausand (5)): Let T_n be the time between the $n-1$ 'th and the n 'th repair, and let $F_n(t) = P(T_n \leq t)$ be the corresponding distribution function. If we assume that the T_n are independently but not equally distributed, the sequence $\{T_n\}_{n \geq 1}$ is called a non-stationary renewal process.

A stationary renewal process is a process that has the same distribution between repairs, whereas the model we use might have different distribu-

tions for the successive times between repairs. At this point the question we have to answer is what distribution is suitable to model the time between repairs, that is, what is the distribution of T_n .

We will estimate the distribution between repair times without making any assumptions, that is we will carry out a non-parametric estimation of F_n . The standard approach for censored data goes via the Kaplan-Meier estimator of the survival function $R_n(t) = P(T_n > t)$. This gives us an estimator \hat{F}_n of the cumulative distribution function F_n as follows

$$\hat{F}_n(t) = 1 - \hat{R}_n(t).$$

The Kaplan-Meier estimation of R_n is

$$\hat{R}_n(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{Y(t_i)}\right), \quad (1)$$

where $Y(t_i)$ is the number at risk just before time t_i and d_i is the number of failures at time t_i , see Høyland and Rausand (5) or Klein and Moeschberger (9). The estimator \hat{F}_n is a step function. Note that when there are no censoring events, \hat{F}_n is the empirical cumulative distribution function.

In order to get an estimate of the probability density function f_n of T_n we can use a kernel estimator

$$\hat{f}_n(t) = \frac{1}{b} \sum_{i=1}^n K\left(\frac{t - t_i}{b}\right) \Delta \hat{F}_n(t_i), \quad (2)$$

where K is a kernel function and b is the bandwidth, see Appendix A. The kernel smoothing function used here is

$$K(x) = \frac{15}{16}(1 - x^2)^2, \quad x \in [-1, 1], \quad (3)$$

but others may also work well.

A kernel estimate can be informative. If there is no prior knowledge about the distribution, a non-parametric estimate is good as a visual confirmation of the choice of parametric distribution. The advantage of a parametric model is that we automatically extrapolate and get knowledge about the distribution outside the field where we observed data. The parameters may sometimes also have a physical meaning. It is more easy to transfer knowledge from RM8 to RM12 if the model is parametric.

Here the Weibull distribution is used with the following parameterization,

$$F(t) = 1 - e^{-(\frac{1}{\theta}t)^\alpha}, \quad t > 0, \quad (\theta > 0, \alpha > 0), \quad (4)$$

where θ is the characteristic life and α is the shape parameter. Let n be the number of repairs we are modeling. Then each time between repairs

is modeled as a Weibull distributed random variable with its own parameters $\{\alpha_i, \theta_i\}$, so that the parameters to be estimated are $\{\theta_1, \dots, \theta_n\}$ and $\{\alpha_1, \dots, \alpha_n\}$. In a later section we will place restrictions on these parameters.

The parameters were estimated by means of the method of maximum likelihood and the method of least sum of squares, see appendix C and D. To see if the Weibull distribution is a reasonable assumption and get an idea of what estimation method to prefer, density functions from all three estimation methods are plotted in Figure 2.

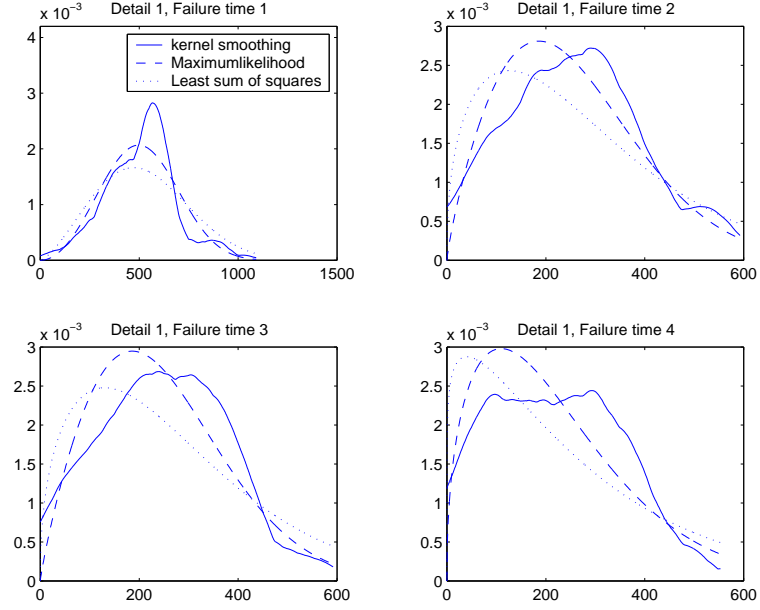


Figure 2: *Maximum likelihood, Least Square and a non-parametric estimation of the density function.*

In Figure 2 kernel functions can be seen as the estimates that are closer to current data. The Maximum likelihood estimate appears closer to the non-parametric estimate than the Least Square estimate. There exist several other formal ways to reach a conclusion, for example the One-sample Test, which with one particular choice of weight function yields the One-sample log-rank test, see Klein and Moeschberger (9). Another visual way of showing if the choice of parametric distribution function is satisfactory is to plot the function $F^{-1}(\hat{F}(t))$ where \hat{F} is the empirical cumulative distri-

bution function from the Kaplan Meier estimation and F is the parametric cumulative distribution function. If F is a good approximation to the data, this should be a straight line. In Figure 3 we see the $F^{-1}(\hat{F}(t))$ function for the four first repair times, with F equal to the Weibull distribution function.

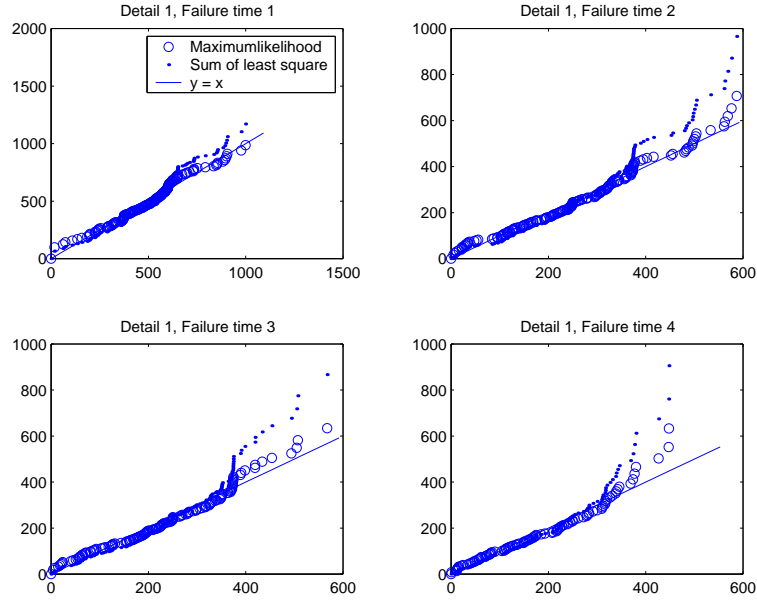


Figure 3: *The $F^{-1}(\hat{F}(t))$ function. If the estimation is good the curve should be close to the straight line $y=x$.*

According to this figure Weibull looks like a reasonable approximation for this data set. We see that the Maximum likelihood estimation is closer to the straight line and hence a better approximation. That the maximum likelihood estimator is a better estimator is also indicated in Beretta and Murakami (2). The choice of the Weibull distribution is not obvious. Other distributions could give a similar approximation of the empirical distribution. In this case the body of the distribution is more important than the tails so the choice of distribution is not that important. Another good thing about the Weibull distribution is that it is well known in industry.

2.3 NHPP-model

Assume that N is a counting process that counts the number of repairs of a component and let the intensity (or hazard) function w be a function of time. If the process has independent and Poisson distributed increments, N is called a Poisson Process. If w is non-constant the process is non-homogeneous. If the number of repairs is Poisson distributed then N is called a non-homogeneous Poisson Process (NHPP).

If this process is used, there will be dependence between repair times. In reality, modeling with NHPP is the same as assuming minimal repair or imperfect repair. That means that when a component is repaired it is repaired to the condition just before the failure occurred. More complicated assumptions can be modeled, see Pham and Wang (11) and Valdez-Flores and Feldman (16). See Høyland and Rausand (5) for a more theoretical description of Poisson processes and applications.

In this model we can estimate the cumulative intensity directly from the data set. The cumulative intensity is defined as

$$W(t) = \int_0^t w(u)du,$$

and W is estimated by the Nelson-Aalen estimator

$$\hat{W}(t) = \sum_{t_i \leq t} \frac{1}{Y(t_i)}, \quad (5)$$

where $Y(t_i)$ is the number of components at risk just before t_i . The times t_i are all times when failures occur. The distribution of T_1 , the time to the first failure is given by

$$P(T_1 > t) = P(N(t) = 0) = e^{-W(t)} = e^{-\int_0^t w(u)du},$$

and the distribution of the time T_i to the next failure given a failure at $T_{i-1} = t_{i-1}$ is given by

$$\begin{aligned} P(T_i > t \mid T_{i-1} = t_{i-1}) &= P(N(s_{i-1} + t) - N(s_{i-1}) = 0) = \\ &= e^{-(W(s_{i-1}+t) - W(s_{i-1}))} = e^{-\int_{s_{i-1}}^{s_{i-1}+t} w(u)du}, \end{aligned}$$

where $s_j = \sum_{k=1}^j t_k$ is the absolute time. The function W represents the mean number of repairs for one component until time t . Estimators of these distributions can be obtained by replacing W with \hat{W} , these are related to the Kaplan-Meier estimator but not identical.

Using a kernel smoother we can estimate the intensity function w itself, the quantity is also known as the hazard rate, see Klein and Moeschberger (9). In Figure 4 estimates of $W(t)$ and $w(t)$ are shown for detail 1, 2 and 3.

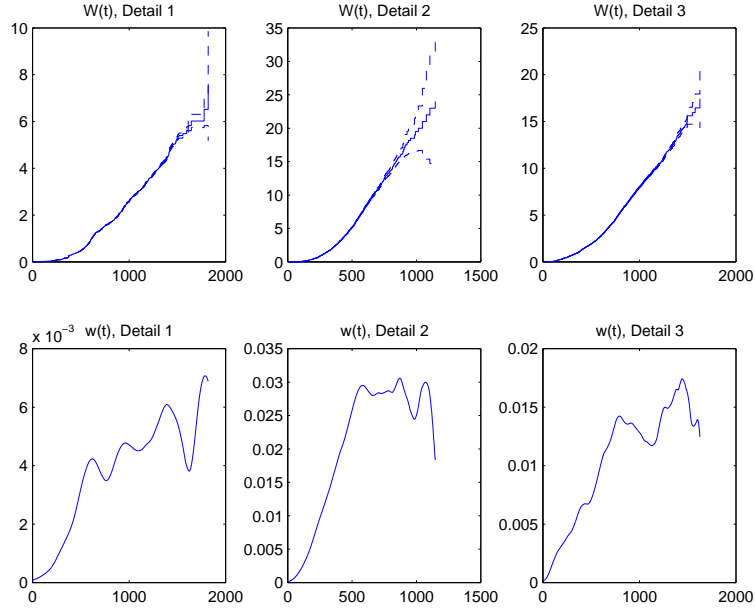


Figure 4: *The Nelson Aalen estimate of W with 90% confidence intervals and kernel estimate of w with kernel function as in (3).*

2.4 Times between failures independent?

When choosing a model it is interesting to know if the times between failures are independent or not. Are the times between repairs related to each other? If the time to first repair comes early is there a greater chance that the component holds longer next time or will the component break early again? Is there any sort of dependency between times of repairs? If the repair times are dependent the NHPP is probably better than the NSRP. In the NSRP the result in the first distribution is independent of the result in the second distribution.

2.5 Which model is the best one?

To decide which model fits data best we must construct some sort of error measurement. There are plenty of measures of fit that are reasonable, e.g.

$$\sup_t | \hat{F}_n(t) - F(t) |, \quad (6)$$

$$\int (\hat{F}_n(t) - F(t))^2 dt, \quad (7)$$

$$\int | \hat{F}_n(t) - F(t) | dt, \quad (8)$$

$$\sum_{i=1}^n (E[T_i] - t_i)^2, \quad (9)$$

$$\sum_{i=1}^n | E[T_i] - t_i |, \quad (10)$$

where $\hat{F}_n(t)$ is the distribution function according to the model, and $F(t)$ is the true distribution function, $E[T_i]$ is the expectation of the model and t_i is the observation. Measure (6) focuses on the greatest difference between the model and the data, (7) and (8) are measures of the overall fit. Measure (9) is a measure of difference in squared mean of the expected outcome and the real outcome, measure (10) is similar to (9) but without square difference penalty. Which measure is the best choice depends on what question we want to answer. In this case we want to use the model to make a prediction of the time when the component fails. As prediction we use the expected failure time according to the model. We want the difference between our prediction and the outcome to be as small as possible. We want a big difference to be more than linear worse than a small difference. This makes it natural to pick (9) as our error measure.

If T_i is the stochastic variable that describes the time to failure, then the error measure m_i for a failure is the quadratic loss function

$$m_i = (E[T_i] - t_i)^2, \quad (11)$$

where t_i is the observed failure time. Another choice that must be made when we compare the models is when to calculate m_i . One way to do this is to calculate $E[T_i]$ directly after component i has failed. This is what we are doing here. In principle it would be possible to calculate this measure at other times as well. For example every 50 hours the component is alive we could measure the time to expected failure according to the model and compare it to the observed failure. Finally we take the mean of all components to get the mean square error

$$M = \frac{1}{n} \sum_{i=1}^n m_i.$$

NSRP: It is straightforward to calculate m_i in the NSRP model since T_i is Weibull distributed and $E[T_i]$ then has a known parametric form. With the parameterization as in (4) we have

$$E[T_i] = \theta_i \cdot \Gamma\left(\frac{1}{\alpha_i} + 1\right),$$

where Γ is the gamma function. The expected squared error in the Weibull distribution is

$$M_e = E[m_i] = E[(E[T_i] - T_i)^2] = V[T_i] = \theta^2 \cdot (\Gamma(\frac{2}{\alpha} + 1) - \Gamma^2(\frac{1}{\alpha} + 1)). \quad (12)$$

The reason why the expected error happens to coincide with the variance of the Weibull distribution is a consequence of the fact that we chose the error measure as we did. Table 1 shows the observed and expected squared error for the four first repairs of detail 1,2 and 3. Note that the total error

detail	parameter	failure 1	failure 2	failure 3	failure 4	Total
1	$M * 10^4$	3.28	1.92	1.57	1.61	2.12
1	$M_e * 10^4$	3.41	2.12	1.89	2.50	2.45
2	$M * 10^4$	2.02	0.06	0.06	0.06	0.34
2	$M_e * 10^4$	2.02	0.06	0.06	0.06	0.34
3	$M * 10^4$	7.20	0.22	0.29	0.28	1.05
3	$M_e * 10^4$	7.45	0.21	0.26	0.26	1.08

Table 1: Calculated error in prediction of the NSRP model for detail 1,2 and 3 for the four first and total time to failure.

here is the mean of more errors than in the four first failure times. It is the mean of all errors to the last failure. As we can see M_e is mostly bigger than M . This may be a consequence of the fact that the true distribution has shorter tails than the Weibull distribution. We can also see the lack of fit in the tail area in Figure 3.

NHPP: In the NHPP the prediction at time t_0 of the time to the next failure T is computed by

$$E[T] = \int_0^\infty e^{-(W(t_0+t) - W(t_0))} dt.$$

Because we do not have any estimation of W for large values of t this is impossible to calculate. The standard way of doing this is to stop integration at the largest observation. Here we present an alternative method of solving this. This method is exact if W is linear and an approximation if W is

close to linear. An advantage of this method is that it gives the possibility to check if the NHPP model is appropriate. If we compare the alternative method of calculating with the standard method the total error became 7% bigger with the alternative way for detail 1 but 4% smaller for detail 2 and 3. The idea behind the alternative approach is to transform the NHPP to a HPP with intensity one. This transformation is given by $\hat{t} = W(t)$ where t is the real time and \hat{t} is the transformed time, c.f. Høyland and Rausand (5). In the HPP the times between errors are independently and exponentially distributed with intensity one. This results in the expected time to next failure being one in the transformed time. We calculate $E[T | T > t_0]$ by means of

1. transform t_0 to \hat{t}_0 ,
2. calculate $E[\hat{T} | \hat{T} > \hat{t}_0] = (\widehat{t_0 + 1})$,
3. transform back $(\widehat{t_0 + 1})$.

We use the alternative way and calculate M for the four first failures. Results are shown in table 2.

detail	parameter	failure 1	failure 2	failure 3	failure 4	Total
1	$M * 10^4$	4.71	1.89	1.65	1.43	2.39
2	$M * 10^4$	2.23	0.42	0.28	0.16	0.43
3	$M * 10^4$	9.77	0.94	0.88	0.72	1.60

Table 2: Calculated error in prediction of the NSRP model for detail 1,2 and 3 in the four first and total time to failure.

If the alternative method is used it is possible to visually control the fact that times between failures are exponentially distributed. Plot the function $F^{-1}(\hat{F}(t))$ where F is the cumulative distribution function of the exponential distribution with intensity one and $\hat{F}(t)$ is the distribution of transformed times between failures. In the data set analysed this is really not the case, especially not for detail 1, as shown in Figure 5. In the HPP, times between failures should be independently distributed if the NHPP is a good model. This can be checked by calculating the correlation between the repair times. The times between repairs are not independently distributed in the data sets examined. This observation indicates that the NHPP-model is not suitable for modeling the behavior of these components.

Comparison A comparison of failure measure between the two models shows that the NSRP model is better modeling this dataset. See Table 3.

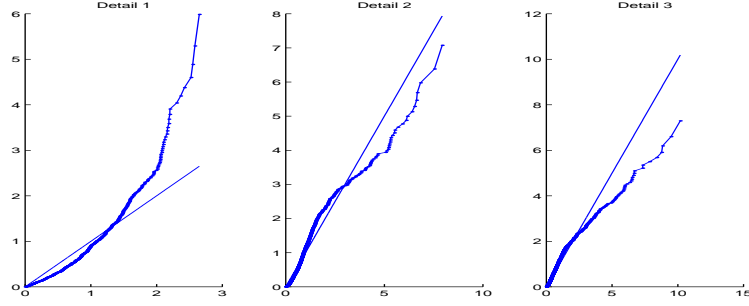


Figure 5: The $F^{-1}(\hat{F}(t))$ function. If the estimation is good the plot follows the line $y=x$.

detail	NSRPM * 10^4	NHPPM * 10^4	$\frac{NHPP_M}{NSRP_M}$
1	2.12	2.39	1.13
2	0.34	0.43	1.26
3	1.05	1.60	1.48

Table 3: Calculated quotient of error measure in the two models

3 Development of the model

Of the two models presented, the NSRP is the better for this type of data. Remember that the main reason for calculating failure times of components for the older engine is that this will help us to understand how similar components of the newer engine will behave. In this section we will see if we can develop the model. A further investigation showed that there are two classes of repairs. We will investigate if these two repairs influence the life of the components differently.

In Figure 2 we can see that the failure distributions seem to be very similar after the first repair. We will use this to refine the model. In Figure 2 we also notice that the time to failure after first repair seemed much longer than the times to the following repairs. This can also be seen in Table 4, if we look at the θ parameter, the characteristic life.

3.1 Different repair stations

There are two different repair stations that repair the details: repair stations within close range of the aircraft A and the main central repair station B. There is reason to believe that the main repair station repairs better than

the smaller repair stations.

To investigate if there is a difference between A and B without making any assumptions we estimate the mean time to repair, μ . Let \hat{R} be the Kaplan Meier estimation (1) of the survival function. The mean μ can be estimated by

$$\hat{\mu} = \int_0^{\tau} \hat{R}(t) dt, \quad (13)$$

where τ is the biggest observed time. The variance of this estimator is

$$V[\hat{\mu}] = \sum_{i=1}^N \left[\int_{t_i}^{\tau} \hat{R}(t) dt \right]^2 \frac{d_i}{Y_i(Y_i - d_i)}, \quad (14)$$

where N is the number of observations, Y_i the number at risk in time t_i and d_i the number of failures at time t_i . A $100(1 - \alpha)$ confidence interval for μ is expressed by

$$\hat{\mu} \pm Z_{1-\frac{\alpha}{2}} \sqrt{V[\hat{\mu}]}. \quad (15)$$

The process $\hat{R}(t)$ is asymptotically normal distributed so the distribution of $\hat{\mu}$ is also asymptotically normal distributed. Theoretical results can be found in Andersen et al. (1).

We estimate μ and calculate the confidence interval according to (13), (14) and (15) but distinguish between repair A and B. In Figure 6 this is shown for the first five repairs with 95% confidence interval for the mean. Note that new details are not shown in this figure. (Repair type B is displayed as a dot and repair type A as a ring. The dots and rings indicate upper and lower confidence bounds. There is also a dot or ring in the middle of the confidence region indicating the point estimation of the mean.) We can see in the figure that there is a difference between the repair A and B, at least for details 1 and 3. The non-parametric estimate $\hat{\mu}$ is shown in Table 4, page 16. There it is possible to compare this mean with the mean achieved if the Weibull distribution was assumed. In Table 4 we see that the means are very close, which is an indication that the Weibull distribution is a good approximation.

3.2 Simplify the model

Now when we are aware of the difference in survival depending on where the component has been repaired we model time to next failure with different parameters depending on where the component has been repaired, but we still use different failure distribution after every repair time, that is F_n depends on n . In Table 4 the θ and α parameters are shown for the time to first failure and for the following five times to failure after repair for detail

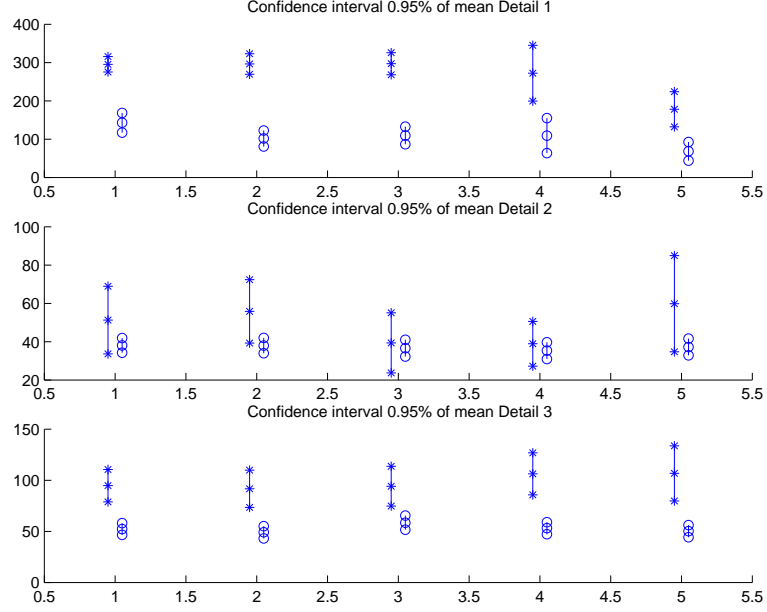


Figure 6: *Confidence interval on expected time to failure after repair number 1 to 5. Dots are repair type B and rings are repair type A.*

1,2 and 3. Also μ_w , the mean of Weibull distribution, and μ , the mean of non-parametric distribution, are shown.

We notice that the estimates of θ and α do not change very much between successive repairs if we disregard new components. This makes it natural to suggest a model with the same distribution for T_i , $i \geq 2$. Maximum likelihood estimates of θ and α are shown in Table 5. Note that these values are close to those in Table 4.

When the details in the new engine design begin in service little is known so this may be a good model to start with. Bayesian updating principles may be usable as well, see e.g. Shimi and Tsokos (14), survey or e.g. Michael and Giuntini (4), Weibull example.

3.3 Aging

An interesting question to ask is if the time to next repair decreases with the number of repairs. If we look at the points in Figure 6, do we see a downward trend? A simple test can be made to answer this question. We

detail	failure nr	rep type	θ	α	μ_w	μ
1	1	-	554	3.09	496	498
1	2	A	145	1.48	131	131
1	3	A	159	1.46	144	142
1	4	A	109	1.26	101	102
1	5	A	125	1.98	110	110
1	6	A	119	1.37	109	109
1	2	B	359	3.00	321	323
1	3	B	330	3.14	295	295
1	4	B	335	2.83	298	296
1	5	B	331	3.69	299	297
1	6	B	297	3.31	266	272
2	1	-	355	2.36	315	315
2	2	A	45	1.76	40	40
2	3	A	43	1.69	38	38
2	4	A	43	1.76	38	38
2	5	A	41	1.69	37	37
2	6	A	40	1.76	36	35
2	2	B	41	2.20	36	36
2	3	B	52	1.04	51	51
2	4	B	62	1.64	55	56
2	5	B	37	0.84	41	39
2	6	B	44	2.29	39	39
3	1	-	589	1.99	522	526
3	2	A	76	1.77	67	67
3	3	A	58	1.61	52	52
3	4	A	54	1.43	49	49
3	5	A	65	1.49	58	58
3	6	A	59	1.66	53	53
3	2	B	101	1.91	90	90
3	3	B	105	1.47	95	95
3	4	B	98	1.27	91	92
3	5	B	103	1.39	94	94
3	6	B	119	1.68	106	106

Table 4: Parameters θ and α in the Weibull distribution, μ_w the mean in this Weibull distribution and μ the non-parametric estimated mean.

test the hypotheses that all μ are equal versus that they are not. To carry out this test we must know the numbers of degrees of freedom. This is a complex thing to find out when we have censored data. If we assume

detail	repair type	θ	α
1	A	134	1.41
1	B	341	2.98
2	A	41	1.77
2	B	49	1.36
3	A	59	1.51
3	B	107	1.55

Table 5: Parameters θ and α in the Weibull distribution if all T_i , $i \geq 2$ were considered to belong to the same distribution.

that the $\hat{\mu}$ is estimated with many observations the estimated variance of $\hat{\mu}$ is near the true variance. We also assume that the variance of each $\hat{\mu}$ is equal. We can then perform a ordinary χ^2 test, described in appendix B. The results turn out to be that we can only reject the hypothesis in the case detail 3 repair type A. Note that this test only includes the five first repairs. If we only look at the estimated values of $\hat{\mu}$ in Figure 6 we may think that there is a larger difference in μ in detail 2 repair type B, but the confidence interval is much bigger, which indicates that we are less certain of the true value.

Another approach to the aging problem is to assume that the deteriorate a little every time they are repaired. If they do not, they may in theory be repaired ad infinitum and still have the same failure distribution. We suggest the following Weibull model for the T_i , $i \geq 2$

$$F(t) = 1 - e^{-(\frac{1}{\theta p^n} t)^\alpha}, \quad t > 0, \quad (\theta > 0, \alpha > 0, p > 0), \quad (16)$$

where n is the repair number. This means that the expected time to failure after repair number n is

$$E[T_n] = \theta_i p^n \cdot \Gamma(\frac{1}{\alpha_i} + 1),$$

and $p < 1$ thus indicates aging. Maximum likelihood estimates of the parameters (θ, α, p) are shown in Table 6.

Here some \hat{p} are bigger than 1. This is an indication that this may not be a good model. If we make a 95% confidence interval over the true parameter p based on profile likelihood we see that in all cases except one we can not reject that $p = 1$. In the case where we could reject $p = 1$, p is very close to one. This means that if we use $p = 1$ the resulting error is very small. No aging parameter is necessary. If the aging parameter is admitted the model gets one extra parameter that will complicate the model but add little new information. Profile likelihood is shortly described in appendix E, more can be found in Venzon and Moolgavkar (17).

detail	repair type	θ	α	p	95% confidence-interval
1	A	164	1.43	0.92	(0.8412 , 1.0146)
1	B	371	3.04	0.95	(0.9205 , 1.0026)
2	A	45.7	1.79	0.98	(0.9666 , 0.9902)
2	B	47.0	1.36	1.01	(0.9205 , 1.0026)
3	A	61.8	1.51	0.99	(0.9794 , 1.0058)
3	B	101.9	1.55	1.01	(0.9846 , 1.0426)

Table 6: Parameters in modified Weibull distribution if all T_i , $i \geq 2$ were considered to belong to the same distribution and a 95% confidence intervals over the parameter p .

4 Discussion and conclusion

We have used two models to predict the time between failures on a data set containing failure times of components in an aircraft engine. The model used was a Non-Stationary Renewal Process (NSRP) and a Non-Homogeneous Poisson Process (NHPP). A question arises with model is preferably to use.

In order to understand and measure what model was most suitable to model current data a couple of error measures were considered. The use of the NSRP model was better for this dataset, see Table 3. In the NSRP model non-parametric estimators as Kaplan-Meier (1) and Nelson Aalen (5) were used at an early stage to avoid making any assumptions on parametric distributions. Kernel smoothing (2) was used to analyse times between successive repair times. With the help of the non-parametric estimations a parametric distribution was chosen. Several plots, Figure 2 and 3, showed that the Weibull distribution modeled the data set well enough to make reasonable predictions. For the choice of error measure (11) the expected error in the NSRP model is equal to the variance of the Weibull distribution (12). However, the error was somewhat smaller, which can be a consequence of the fact that the real distribution has smaller tails than the Weibull distribution.

Looking closer at the model, we found that different repair stations had an impact on the times between repairs, see Table 6. We would like to model the time to next failure with different distributions depending on where the component has been repaired. We also noted that the first time to failure was much longer than the following times. However, the following times were from the same distributions. A more advanced model that suggested that the components deteriorated with time (16) was suggested but it was rejected by data.

The model we suggest has a distribution to the first failure F_{new} and

then two different distributions to the following failures F_A and F_B depending on where the component was repaired. We have 6 parameters to estimate, $\{\theta_{new}, \alpha_{new}, \theta_A, \alpha_A, \theta_B, \alpha_B\}$ see Table 5. We use $\{\theta_{new}, \alpha_{new}\}$ to model the time to next failure if the component is new and $\{\theta_A, \alpha_A\}$ if the component has been repaired at station A, and finally we model with $\{\theta_B, \alpha_B\}$ if the component has been repaired at station B.

The models used data from an older version of an aircraft engine called RM8. A newer engine called RM12 is available but little data exist for that engine. The models in this work can of course be used for RM8 but the main goal was to use them for RM12. That is one of the reasons why a parametric model was chosen since a parametric model may be easier to transfer to the RM12 case than a non-parametric model. However, more work needs to be done in this area.

Acknowledgements

The author wants to thank Dragi Anevski and Thomas Svensson at Fraunhofer Chalmers Center and Jacques de Maré at Chalmers for supervision. Furthermore I own a debt of gratitude to Fredrik Plym and Maud Österman at Volvo Aero Corporation for data support. This work was financed by NFFP.

Appendix

A: Kernel smoothing; Let $H(t)$ be a step function with jumps at the event times $t_1 < t_2 < \dots < t_n$. Let $\Delta H(t_i) = H(t_i) - H(t_{i-1})$ denote the magnitude of the jumps in $H(t_i)$ at time t_i . The kernel smoothed estimator of $h(t)$ is a weighted average of values of $\Delta H(t_i)$ for t_i close to t . Closeness is determined by a bandwidth b so that $t_i \in [t - b, t + b]$ are included in the weighted average. The bandwidth is chosen either to minimise some measure or to give a desired degree of smoothness. Let $K(t)$ be the kernel function that describes how much weight is given to points at a distance from t . Three common kernel functions for $x \in [-1, 1]$ are

$$K(x) = \frac{1}{2}; \quad K(x) = \frac{3}{4}(1 - x^2); \quad K(x) = \frac{15}{16}(1 - x^2)^2.$$

The estimation is given by

$$\hat{h}(t) = \frac{1}{b} \sum_{i=1}^n K\left(\frac{t - t_i}{b}\right) \Delta H(t_i).$$

When $t > t_n - b$ and $t < t_1 + b$ this estimate is biased but can be corrected. More information about kernel smoothing can be found in Klein and Moeschberger (9).

B: χ^2 test; We have 5 estimated means $\{\hat{\mu}_1, \dots, \hat{\mu}_5\}$, each with an estimated variance $\{S_1^2, \dots, S_5^2\}$. We want to test $H_0 : \mu_i$ equal vs $H_1 : \mu_i$ not equal. The number of observations we used to estimate μ_i varies because we use censored data. We assume that μ_i is estimated with a great deal of data and hence S_i^2 is near the true variance. We also assume that all the variances are equal. We calculate

$$S_w = \frac{1}{5} \sum_{i=1}^5 S_i^2; \quad S_b = \frac{1}{4} \sum_{i=1}^5 (\hat{\mu}_i - \bar{\mu})^2; \quad \text{where } \bar{\mu} = \frac{1}{5} \sum_{i=1}^5 \hat{\mu}_i.$$

and

$$\frac{4S_b}{S_w} \sim \chi_4^2.$$

The P-values of this test is shown in Table 7. We can only reject the H_0

detail	repair type	P-values
1	A	0.245
1	B	0.530
2	A	0.226
2	B	0.599
3	A	0.001
3	B	0.712

Table 7: P-values of the test $H_0 : \mu_i$ equal vs $H_1 : \mu_i$ not equal.

hypothesis in the detail 3 repair type A case.

C: Maximum likelihood estimator; We observe $\{t_i, \dots, t_r\}$ as failures and $\{t_{r+1}, \dots, t_n\}$ as censored times. We want to estimate the parameters in (4) by the maximum likelihood method. If the censoring process is random the maximum likelihood estimator is

$$L(\theta, \alpha | t_i) = \prod_{i=1}^r f(\theta, \alpha | t_i) \cdot \prod_{i=r+1}^n R(\theta, \alpha | t_i), \quad (17)$$

and the maximum likelihood estimation, $\hat{\theta}$ and $\hat{\alpha}$ of θ and α is obtained by

$$(\hat{\theta}, \hat{\alpha}) = \underset{\theta, \alpha}{\operatorname{argmax}} L(\theta, \alpha | t_i).$$

D: Least Square Method; To estimate parameters in (4) by the least square method we rewrite F on a form that is linear in the parameters,

$$\log(-\log(1 - F(t))) = \log\left(\frac{t}{\theta}\right)^\alpha = \alpha \log(t) - \alpha \log(\theta). \quad (18)$$

We handle the censored observations by estimate $R(t) = 1 - F(t)$ with Kaplan-Meier estimator (1). We denote

$$\mathbf{Y} = \begin{pmatrix} \log(-\log(\hat{R}(t_1))) \\ \vdots \\ \log(-\log(\hat{R}(t_n))) \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} 1 & \log(t_1) \\ \vdots & \vdots \\ 1 & \log(t_r) \end{pmatrix}; \quad \mathbf{P} = \begin{pmatrix} a \\ b \end{pmatrix}$$

where

$$-\alpha \log(\theta) = a; \quad -\alpha = b$$

and consider the equations $\mathbf{Y} = \mathbf{XP}$. Let $\hat{\mathbf{P}}$ be the estimation of \mathbf{P} and $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{P}}$ the expected failures under $\hat{\mathbf{P}}$. The vector $\hat{\mathbf{P}}$ that minimises $(\mathbf{Y} - \hat{\mathbf{Y}})^T \cdot (\mathbf{Y} - \hat{\mathbf{Y}})$ is called the least square estimation of \mathbf{P} and is calculated by

$$\hat{\mathbf{P}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (19)$$

The parameters θ and α are obtained by

$$\alpha = -b \quad \text{and} \quad \theta = e^{-\frac{a}{b}}$$

One could argue that t_i is the random component and not $\hat{R}(t_i)$. In that case we rewrite (18) to

$$\frac{\log(-\log(R(t)))}{\alpha} + \log(\theta) = \log(t)$$

and define

$$\mathbf{Y} = \begin{pmatrix} \log(t_i) \\ \vdots \\ \log(t_n) \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} 1 & \log(-\log(\hat{R}(t_1))) \\ \vdots & \vdots \\ 1 & \log(-\log(\hat{R}(t_n))) \end{pmatrix}; \quad \mathbf{P} = \begin{pmatrix} a \\ b \end{pmatrix}$$

where

$$\log(\theta) = a; \quad \frac{1}{\alpha} = b.$$

We may now minimise $(\mathbf{Y} - \hat{\mathbf{Y}})^T \cdot (\mathbf{Y} - \hat{\mathbf{Y}})$ by (19) and obtain θ and α by

$$\alpha = \frac{1}{b} \quad \text{and} \quad \theta = e^a.$$

We have used the first approach in this report.

E: Profile Likelihood; To make a confidence interval of the parameter p in (16) we first define the profile likelihood

$$\tilde{L}(p) = \max_{\alpha, \theta} L(p, \alpha, \theta),$$

where L is defined in (17) and let

$$\tilde{l}(p) = \log(\tilde{L}(p)),$$

be the log-likelihood function. Let

$$\hat{p} = \operatorname{argmax}_p \tilde{l}(p).$$

Then

$$2(\tilde{l}(\hat{p}) - \tilde{l}(p)) \sim \chi_2^2,$$

and a 95% confidence region is

$$\tilde{R}_c = \left\{ p : \tilde{l}(\hat{p}) - \tilde{l}(p) \leq \frac{\chi_2^2(0.95)}{2} \right\}.$$

References

- [1] P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, 1993.
- [2] S. Beretta and Y. Murakami. Statistical analysis of defects for fatigue strength prediction and quality control of materials. *Fatigue Fract. Engng. Mater. Struct.*, 21:1049–1065, 1998.
- [3] Emad El-Newehi and Frank Proschan. Multistate reliability models: a survey. In *Multivariate analysis, V (Proc. Fifth Internat. Sympos., Univ. Pittsburgh, Pittsburgh, Pa., 1978)*, pages 523–541. North-Holland, Amsterdam, 1980.
- [4] Michael E. Giuntini, Ronald E.; Giuntini. Simulating a weibull posterior using bayes inference. *Annual Reliability and Maintainability Symposium*, pages 48–55, 1993.
- [5] A. Høyland and R. Rausand. *System Reliability Theory*. John Wiley & Sons Inc., 1994.
- [6] J. Kalbfleisch and R. Prentice. *The statistical analysis of failure time data*. John Wiley and sons, New York, 1980.

- [7] V.V. Kaminskiy, M.P.; Krivtsov. G-renewal process as a model for statistical warranty claim prediction. *Annual Reliability and Maintainability Symposium*, pages 276–280, 2000.
- [8] Masaaki Kijima. Generalized renewal processes and general repair models. In *Stochastic models in reliability and maintenance*, pages 145–164. Springer, Berlin, 2002.
- [9] J. Klein and M. Moeschberger. *Survival Analysis*. Springer-Verlag Inc., New York, 1997.
- [10] M. J. Krok and M. J. Ashby. Condition-based diagnostic gas path reasoning for gas turbine engines. In *Proceeding of the 2002 IEEE International Conference on Control Applications*, pages 227–234. 2002.
- [11] H. Pham and H. Wang. Imperfect maintenance. *European Journal of Operational Research*, 94:425–438, 1996.
- [12] M.J. Roemer and D.M. Ghiocel. Probabilistic approach to the diagnosis of gas turbine engines. Technical report, 53rd Machinery Prevention Technologies (MFPT), Conference, Virginia Beach, VA, April 1999.
- [13] M.J. Roemer and G.J. Kacprzyński. Advanced diagnostics and prognostics for gas turbine engine risk assessment. Technical report, IGTI/ASME Turbo Expo, Munich, Germany, May 2000.
- [14] I. Shimi and C. P. Tsokos. The Bayesian and nonparametric approach to reliability studies: a survey of recent work. In *The theory and applications of reliability, with emphasis on Bayesian and nonparametric methods (Conf., Univ. South Florida, Tampa, Fla., 1975), Vol. I*, pages 5–47. Academic Press, New York, 1977.
- [15] W.P.J. Visser T. Tinga and W.B de Wolf. Integrated lifing analysis for gas turbine components. Technical Report NLR-TP-2000-632, National Aerospace Laboratory NLR, November 2000.
- [16] C. Valdez-Flores and R.M. Fieldman. A survey of preventive maintenance models for stochastically deteriorating single-unit systems. *Naval Research Logistics* 36, 419–446, 36:419–446, 1989.
- [17] D.J Venzon and S.H Moolgavkar. A method for computing profile-likelihood-based confidence intervals. *Appl. Statist.*, 37:87–94, 1988.
- [18] George H. Weiss. A survey of some mathematical models in the theory of reliability. In *Statistical Theory of Reliability (Proc. Advanced Seminar, Math. Res. Center, U.S. Army, Univ. of Wisconsin, Madison,*

Wis., 1962), pages 3–54. Univ. of Wisconsin Press, Madison, Wis., 1963.

- [19] S.D. Yang, J.N; Manning. A simple second order approximation of stochastic crack growth analysis. *Engineering Fracture Mechanics*, 53:677–686, 1996.

Paper B

Discrete Approximations of Life Distributions in Optimal Replacement

Johan Svensson
Chalmers University of Technology
Department of Mathematical Statistics
SE-41296 Göteborg Sweden

Abstract

Discretization of continuous random variables is used in a stochastic optimization problem. We present a measure connected to an optimization model that suggests what parts to replace in an aircraft engine. The optimization model requires a discrete random variable with restrictions on the points of support. Four methods to make discretizations are discussed and adapted to the constraints of the model. The methods are compared and the choice of the number of points of support is discussed. Finally the consequence of using a so called narrow scenario tree is commented upon.

Keywords: optimal maintenance; discretization; points of support

1 Introduction

Aircraft engines can be more economically maintained and resources saved if the maintenance process is optimal. The optimality is here defined in economical terms, and the main factor for economical savings determines which components should be replaced on each service occasion. Several optimization models have been developed to deal with this problem. Epstein and Wilamowsky (4) and Dickman et al. (5) have developed models for components with predetermined deterministic lives. Andréasson (2) has developed a model for details with deterministic lives but also details with stochastic lives. Another approach to the problem is to construct a general maintenance policy that is not always optimal but at least satisfactory. A survey of replacement and maintenance policies can be found in Wang (14).

We will use the model presented in Andréasson (2), formulate it as a two-stage model, and define an error measure. Different discrete approximations will be used in order to incorporate components with stochastic

lives into the model. With the help of the error measure we will compare different approximations. Work on different ways of making discretizations of continuous distribution functions for implementation in models has been done by Høyland and Wallace (7), Keefer (9), Miller and Rice (10) and Smith (12), (13). Other work on discretization and tree reduction in multi-stage problems has been done by i.e Dupačová et al. (3), Frauendorfer and Schürle (6) and Pflug (11). We will make discretizations with restrictions on the points of support of the density function.

This paper consists of three parts. In the first part we present the optimization model and the error measure. The second part introduces different discretizations methods, and the last part compares the different methods with respect to the error measure. The objective is to conclude how many points of support are necessary and what method of discretization to use. Furthermore, we discuss the consequence of making a narrow scenario tree.

2 Optimization model

The aim of the optimization model is to minimize the expected cost of maintenance for an aircraft engine. In order to describe the maintenance we classify the components of the engines into two categories; deterministic details and stochastic details.

Definition 1: A *deterministic detail* is a component that has a predetermined limited time in service which must not be exceeded.

Components that are vital for safety are deterministic. If a deterministic component fails there is a risk that the engine will stop functioning. The predetermined time limit is set so low that there is practically no risk the component will fail before this time.

Definition 2: A *stochastic detail* is a component that is not deterministic.

Stochastic components are allowed to operate in the aircraft engine until they fail. If a stochastic component fails during a flight the engine will still work but with reduced performance. In this paper only one stochastic detail is used but it is possible to extend the model to several stochastic details. The life U of a new stochastic component is modeled with a distribution \tilde{G} and the remaining life of a functioning stochastic component with an age

of u_0 is modeled with a distribution G where

$$\begin{aligned} G(u) &= P(U \leq u + u_0 \mid U > u_0) = \frac{P(u_0 < U \leq u + u_0)}{1 - P(U \leq u_0)} \\ &= \frac{\check{G}(u + u_0) - \check{G}(u_0)}{1 - \check{G}(u_0)}. \end{aligned}$$

We assume that $\check{G}'(u) > 0$ if $u > 0$ and U to be a non-negative random variable.

The engine is repaired for two reasons, either the stochastic detail has failed or a deterministic detail has reached its predetermined time limit. The repair of the engine is associated with a cost for bringing the engine to service. At service, there is an opportunity to replace other components and consequently extend the time to the next repair. Every time the engine is at the repair bay, an optimization is performed. The input data is the state of the components in the engine and the output data is a suggestion of what parts to replace. The optimization model is described in detail in Andréasson (2).

In order to solve the optimization problem, for an engine consisting of many parts, simplifications must be made. The main simplification in this paper is that in the sequence of life distributions for the stochastic detail only the first life distribution is modeled in a stochastic way, the remaining life distributions are replaced by the expected value of the life distribution. This simplification makes the model easier to formulate and solve, but there are some drawbacks, described in Section 4.2.

Another simplification inherent in the optimization model is that the time, during which the maintenance cost is minimized, is divided into T discrete time points instead of a continuous approach. The optimization model is formulated as a stochastic two-step model. More information about stochastic optimization models can be found in Kall and Wallace (8).

2.1 First stage model

In order to decide which components to replace we introduce the first step binary variables x_0^i for deterministic details $i = 1, \dots, N$ and \tilde{s}_0 for the stochastic detail. The variables $x_0^1, \dots, x_0^N, \tilde{s}_0$ take the values 0 or 1 with a 1 indicating replacement of the detail. The lower index indicates the time point and zero indicates that this is a decision we have to make now, hence a first step variable. We introduce the replacement strategy vector $\mathbf{x}_0 = (x_0^1, \dots, x_0^N, \tilde{s}_0)$ and we want to find the solution

$$\hat{\mathbf{x}}_0 = \underset{\mathbf{x}_0 \in \{0,1\}^{N+1}}{\operatorname{argmin}} F(\mathbf{x}_0),$$

with

$$F(\mathbf{x}_0) = \int_0^\infty f(\mathbf{x}_0, u) dG(u) = E_G[f(\mathbf{x}_0, U)], \quad (1)$$

where f is defined in section 2.2.

2.2 Second stage model

In the second stage model we take into consideration possible future problems. Overall we want to minimize the expected cost of maintaining the engine during a fix time period containing T equidistant time points. We denote the time points nodes. One approximation of the optimization model is that components are only allowed to break and be replaced at nodes. If T is large we get good resolution at the cost of long calculation time, and if T is smaller the resolution gets worse but the calculation time shortens.

The lives of the deterministic components that are currently in the engine correspond to the node indices $\tilde{\tau}_1, \dots, \tilde{\tau}_N$.

The life of the stochastic component currently in the engine U is transformed to a node by $\tilde{\tau}_s = \tilde{\tau}_s(u)$, which is a function from U to an index set $1, \dots, T$ describing which node corresponds to every u . For later convenience let us construct $\tilde{\tau}_s(u)$ as a step function in such a way that it is right continuous, increasing in u , and does not jump at any nodes. Furthermore if U takes the exact time corresponding to node i , we let $\tilde{\tau}_s(u) = i$. This implies that f (defined in (2) below) for a fixed \mathbf{x}_0 is a right continuous step function in u and does not jump at any node. The reason for this is that f only changes values when $\tilde{\tau}_s(u)$ is changing values. The function f also decreases in u (formulated as a lemma in section 2.4). The function $\tilde{\tau}_s$ can be written as

$$\tilde{\tau}_s(u) = i \text{ if } u \in [w_{i-1}u_{i-1} + (1 - w_{i-1})u_i, w_i u_i + (1 - w_i)u_{i+1}),$$

where u_i is the time corresponding to node i and $w_i \in (0, 1)$. This definition means that we treat components that are going to break some time after u_i as broken at time u_i and we therefore replace them at time u_i . This is an approximation since at time u_i we never know if the components will break in the near future. The motivation for this approach is that we are interested in the first stage variables \mathbf{x}_0 and not the replacement schemes (the second stage variables) if we get the outcome u . Remember here that the nodes do not correspond to any actual maintenance times. In reality a stochastic component is replaced as soon as it is observed as broken, whether this times corresponds to a node or not.

The lives of new (replacing) deterministic and stochastic details are described with node indices τ_i, \dots, τ_N and τ_s with τ_s indicating the node closest to the expected value of the life distribution \tilde{G} .

We introduce the second stage binary variables as $(x_t^1, \dots, x_t^N), t = 1, \dots, T$ for the deterministic components, $\tilde{s}_t, t = 1, \dots, T$ for the stochastic component currently in the engine, and $s_t, t = 1, \dots, T$ for the replacing stochastic components. Furthermore, we let z_t be binary variables indicating whether service is performed on the engine at time $t = 1, \dots, T$. Thus the vectors $(x_t^1, \dots, x_t^N, \tilde{s}_t, s_t, z_t)$ describe if replacement and service are performed at times $t = 1, \dots, T$ with 1 indicating replacement and service, respectively.

Costs associated with the maintenance are c_i , c_s , and d , where c_i is the cost of replacing component $i = 1, \dots, N$, c_s the cost of replacing the stochastic component and d a fixed cost for the service process. Thus $(c_1, \dots, c_N, c_s, d)$ is the vector of costs associated with the replacements and service made at each node.

In order to simplify the notation we introduce the set $\mathcal{N} = \{1, \dots, N\}$. The second stage function f describes the cost if we make replacement \mathbf{x}_0 , when the stochastic life is u and is defined as

$$f(\mathbf{x}_0, u) = \min_{\mathbf{x} \in \psi} \sum_{t=0}^T \left(\sum_{i \in \mathcal{N}} c_i x_t^i + c_s(s_t + \tilde{s}_t) + dz_t \right), \quad (2)$$

where ψ is the set of points $\mathbf{x} = (x_t^1, \dots, x_t^N, s_t, \tilde{s}_t, z_t; t = 1, \dots, T)$ such that

$$x_0^i + \sum_{t=1}^{\tilde{\tau}_i} x_t^i \geq 1, \quad i \in \mathcal{N}, \quad (3)$$

$$\sum_{t=\ell}^{\tau_i+\ell-1} x_t^i \geq 1, \quad \ell = 0, \dots, T - \tau_i, \quad i \in \mathcal{N}, \quad (4)$$

$$x_t^i \leq z_t, \quad t = 0, \dots, T, \quad i \in \mathcal{N}, \quad (5)$$

If $\tilde{\tau}_s(u) \geq T$ constraints (6)-(13) shall be removed

$$\sum_{t=0}^T \tilde{s}_t = 1, \quad (6)$$

$$\sum_{t=0}^{\tilde{\tau}_s} \tilde{s}_t = 1, \quad (7)$$

$$\sum_{t=\ell}^{\tau_s+\ell-1} s_t + \sum_{t=\ell}^{\tilde{\tau}_s(u)} \tilde{s}_t \geq 1, \quad \ell = 0, \dots, \min(\tilde{\tau}_s(u), T - \tau_s), \quad (8)$$

$$\sum_{t=\ell}^{\tau_s+\ell-1} s_t \geq 1, \quad \ell = \tilde{\tau}_s(u) + 1, \dots, T - \tau_s, \quad (9)$$

$$s_0 = 0, \quad (10)$$

$$s_t \leq \sum_{k=0}^{t-1} \tilde{s}_k, \quad t = 1, \dots, \tilde{\tau}_s(u), \quad (11)$$

$$\tilde{s}_t \leq z_t, \quad t = 0, \dots, \tilde{\tau}_s(u), \quad (12)$$

$$s_t \leq z_t, \quad t = 0, \dots, T, \quad (13)$$

$$z_0 = 1, \quad (14)$$

$$x_t^i, s_t, \tilde{s}_t, z_t \in \{0, 1\}, \quad t = 0, \dots, T, \quad i \in \mathcal{N}. \quad (15)$$

The constraints (3) force the installed deterministic components to be replaced before their lives $\tilde{\tau}_i$ are consumed. The components that replace the current deterministic components are not allowed to be in the engine more than τ_i nodes. This is regulated by constraints (4). Constraints (5) force the indicator variables z_t to be one if we replace any deterministic component at node t .

Constraints (6) to (13) are constraints regulating replacement of the stochastic component and should be removed if $\tilde{\tau}_s(u) \geq T$, because then replacement of the stochastic component is not necessary. Constraint (6) tells us that the installed stochastic component can only be replaced once and (7) tells us that the replacement must occur before node $\tilde{\tau}_s$. The component that replaces the current stochastic component is not allowed to be in the engine more than τ_s nodes. This is regulated by constraints (8) and (9).

Constraints (10) and (11) mean that no replacing stochastic component is allowed to be installed before the current stochastic component is removed.

Constraints (12) and (13) force the z_t variables to be one if we replace the stochastic component at node t . The engine is at the repair bay at time zero, hence constraint (14) sets $z_0 = 1$. In constraints (15) we have the binary restrictions.

Note that in this model it is only the time to first failure $\tilde{\tau}_s$ of the stochastic component that is modeled with distribution G . The remaining times are all assumed (modeled) to be deterministic and equal to the node τ_s closest to expectation under \check{G} . This is sometimes called a narrow scenario tree, see Altenstedt (1). In the model there is no requirement that the components function at node T .

2.3 Maximum discretization

In the above formulation $\tilde{\tau}_s$ corresponds to a discretization of G . The structure of the second stage model requires discretization to work. In section 2.4 we will measure how good a discretization is. In order to do that we introduce two kinds of discretizations. The first discretization allows probability mass on all nodes and the second allows probability mass only at a subset of the nodes.

Let $\kappa_T = \{u_1, \dots, u_T\}$ be the maximal set of nodes where probability mass is allowed, let $n < T$ be a positive integer, and denote $\kappa_n = \{k_1, \dots, k_n\} \subset \kappa_T$.

Definition 3: The discretization G_T is a discretization of G that allows probability mass in *all* T nodes in κ_T . We call this discretization the *maximum discretization*.

Definition 4: The discretization G_n is a discretization of G that allows probability mass in at most n nodes in κ_T . We call this discretization the *n-node discretization*.

A maximum discretization G_T has the corresponding probability mass function

$$g_T(u) = \begin{cases} p_{1_T}, & \text{if } u = u_1, \\ \vdots & \vdots \\ p_{T_T}, & \text{if } u = u_T, \end{cases} \quad (16)$$

and a n-node discretization G_n the probability mass function

$$g_n(u) = \begin{cases} p_1, & \text{if } u = k_1, \\ \vdots & \vdots \\ p_n, & \text{if } u = k_n. \end{cases} \quad (17)$$

We discuss different discretization approaches in section 3.

Nodes with non-zero probability mass are also called points of support for a distribution. Points of support for a distribution $H(u)$ are all points u such that, if $a < u < b$, then $H(b) - H(a) > 0$, for any a, b .

2.4 Measure of error

In order to reduce the time needed to complete the optimization, we want a discretization with as few points of support as possible but we still want a good replacement strategy. Using the maximum discretization (16) or

the n-node discretization (17) yields different replacement strategies. The maximum discretization gives the replacement strategy

$$\hat{\mathbf{x}}_0^T = \underset{\mathbf{x}_0 \in \{0,1\}^{N+1}}{\operatorname{argmin}} F_T(\mathbf{x}_0), \quad (18)$$

where $F_T(\mathbf{x}_0)$ is defined in (1) with G replaced by G_T , so that

$$F_T(\mathbf{x}_0) = \sum_{i=1}^T f(\mathbf{x}_0, u_i) p_{i_T}.$$

The n-node discretization (17) gives the replacement strategy

$$\hat{\mathbf{x}}_0^n = \underset{\mathbf{x}_0 \in \{0,1\}^{N+1}}{\operatorname{argmin}} F_n(\mathbf{x}_0), \quad (19)$$

with F_n defined in (1) with G replaced by G_n , so that

$$F_n(\mathbf{x}_0) = \sum_{i=1}^n f(\mathbf{x}_0, u_i) p_i.$$

As G_T is a distribution using the maximum number of nodes in the model we compare the quality of G_n to G_T by introducing the error measure for the expected cost between two discretizations G_T and G_n as

$$e(G_n, G_T) = F_T(\hat{\mathbf{x}}_0^n) - F_T(\hat{\mathbf{x}}_0^T). \quad (20)$$

Note that $e(G_n, G_T) \geq 0$. In order to make a discretization that has a small error we use the following result.

Theorem 1: The error measure can be bounded with the following inequalities,

$$e(G_n, G_T) \leq 2 \sup_{\mathbf{x}_0} |F_n(\mathbf{x}_0) - F_T(\mathbf{x}_0)| \leq C \sup_u |G_n(u) - G_T(u)|, \quad (21)$$

where C is a bounded constant.

The first inequality is proven in Pflug (11) but the short proof is stated here for the convenience of the reader.

Proof:(First inequality) Set $\varepsilon = \sup_{\mathbf{x}_0} |F_n(\mathbf{x}_0) - F_T(\mathbf{x}_0)|$. Let $M = \{\mathbf{x}_0 : F_T(\mathbf{x}_0) \leq F_T(\hat{\mathbf{x}}_0^T) + 2\varepsilon\}$. Suppose that $\hat{\mathbf{x}}_0^n \notin M$, then

$$F_T(\hat{\mathbf{x}}_0^T) + 2\varepsilon < F_T(\hat{\mathbf{x}}_0^n) \leq F_n(\hat{\mathbf{x}}_0^n) + \varepsilon \leq F_n(\hat{\mathbf{x}}_0^T) + \varepsilon \leq F_T(\hat{\mathbf{x}}_0^T) + 2\varepsilon.$$

This contradiction establishes $\hat{\mathbf{x}}_0^n \in M$, i.e.

$$e(G_n, G_T) = F_T(\hat{\mathbf{x}}_0^n) - F_T(\hat{\mathbf{x}}_0^T) \leq 2\varepsilon = 2 \sup_{\mathbf{x}_0} |F_n(\mathbf{x}_0) - F_T(\mathbf{x}_0)|. \square$$

In the second inequality in (21) we use the fact that the second stage function f in (1) is right continuous and does not jump at any nodes. We also use the following lemma.

Lemma 1: The second stage function f decreases in u for any fix \mathbf{x}_0 , i.e

$$f(\mathbf{x}_0, u) \geq f(\mathbf{x}_0, u + \varepsilon), \quad \forall \varepsilon > 0$$

Proof: We will check that the point $\mathbf{x} = (x_t^1, \dots, x_t^N, s_t, \tilde{s}_t, z_t; t = 1, \dots, T)$, which solves the minimization problem (2) when $u = u_0$ still satisfies the constraints when $u = u_0 + \varepsilon$. This means that $f(\mathbf{x}_0, u)$ decreases in u since we can guarantee the same cost at $u = u_0 + \varepsilon$ as when $u = u_0$.

The relevant constraints to consider are (7), (8), (9) and (11). In the case when $\tilde{\tau}_s(u_0) = \tilde{\tau}_s(u_0 + \varepsilon)$, all constraints are the same. In the case when $\tilde{\tau}_s(u_0) < T$ and $\tilde{\tau}_s(u_0 + \varepsilon) \geq T$, \mathbf{x} belongs to ψ when $u = u_0 + \varepsilon$ since removing constraints means less restriction on the second stage variables. We now consider the case when $\tilde{\tau}_s(u_0) < \tilde{\tau}_s(u_0 + \varepsilon) < T$.

We know from constraints (6) and (7) that \mathbf{x} satisfies $\tilde{s}_t = 0$, $t \geq \tilde{\tau}_s(u_0) + 1$. This observation makes (7) true when $u = u_0 + \varepsilon$.

We now consider constraints (8) and (9) when $u = u_0 + \varepsilon$ but $\tilde{s}_t = 0$, $t \geq \tilde{\tau}(u_0) + 1$. The first $\tilde{\tau}_s(u_0) + 1$ constraints in (8) are identical to the constraints in (8) when $u = u_0$. The remaining $\tilde{\tau}(u_0 + \varepsilon) - \tilde{\tau}(u_0)$ constraints in (8) are identical to the first $\tilde{\tau}(u_0 + \varepsilon) - \tilde{\tau}(u_0)$ constraints in (9) when $u = u_0$. All the constraints in (9) when $u = u_0 + \varepsilon$ can be found in (9) when $u = u_0$.

The first $\tilde{\tau}_s(u_0)$ constraints in (11) are identical when $u = u_0$ and $u = u_0 + \varepsilon$. When $u = u_0 + \varepsilon$ and we use $\tilde{s}_t = 0$, $t \geq \tilde{\tau}(u_0) + 1$ the remaining $\tilde{\tau}_s(u_0 + \varepsilon) - \tilde{\tau}_s(u_0)$ constraints state

$$s_t \leq \sum_{i=1}^{\tilde{\tau}_s(u_0)} \tilde{s}_i, \quad t = \tilde{\tau}(u) + 1, \dots, \tilde{\tau}(u + \varepsilon),$$

which is not really a restriction since $\sum_{i=1}^{\tilde{\tau}_s(u_0)} \tilde{s}_i = 1$ according to constraint (7), and all variables are binary according to (15).

We have now checked that solution \mathbf{x} does not violate any constraints when $u = u_0 + \varepsilon$ since the constraints do not mean any restrictions or are identical to the ones occurring when $u = u_0$. \square

A consequence of Lemma 1 is that the longer the life of the stochastic detail is the cheaper it is to maintain the engine.

Proof:(Second inequality): Denote $Y = f(\mathbf{x}_0, U) = f(U)$ and note that Y is a non-negative stochastic variable. We rewrite

$$P(Y > y) = 1 - P(f(U) \leq y) = 1 - P(U \geq f^{-1}(y)) = P(U < f^{-1}(y)),$$

where

$$f^{-1}(y) = \begin{cases} \infty, & \text{if } y < f_{min}, \\ \inf\{z : f(z) \leq y\}, & \text{if } f_{min} \leq y \leq f_{max}, \\ 0, & \text{if } y > f_{max}, \end{cases}$$

and f_{min} and f_{max} depend on \mathbf{x}_0 and are the minimum and maximum costs of maintaining the engine. Note that $P(U < f^{-1}(y)) = G(f^{-1}(y))$ if f does not change values at node points, i.e. if $f^{-1}(y) \notin \kappa_T$. Note that this follows from the assumption on $\tilde{\tau}_s$.

Equation (1) can be rewritten

$$\begin{aligned} E_G[Y] &= \int_0^\infty P(Y > y) dy = \int_0^\infty P(U < f^{-1}(y)) dy \\ &= \int_0^\infty G(f^{-1}(y)) dy = f_{min} + \int_{f_{min}}^{f_{max}} G(f^{-1}(y)) dy. \end{aligned}$$

Now

$$\begin{aligned} & 2 \sup_{\mathbf{x}_0} | F_n(\mathbf{x}_0) - F_T(\mathbf{x}_0) | \\ &= 2 \sup_{\mathbf{x}_0} | E_{G_T}[Y] - E_{G_n}[Y] | \\ &= 2 \sup_{\mathbf{x}_0} \left| \int_{f_{min}}^{f_{max}} G_T(f^{-1}(y)) - G_n(f^{-1}(y)) dy \right| \\ &\leq C \sup_u | G_T(u) - G_n(u) |, \end{aligned}$$

with $C = 2 \sup_{\mathbf{x}_0} (f_{max} - f_{min})$. Then

$$f_{max} \leq \hat{f}_{max} = \min \left\{ \sum_{i \in \mathcal{N}} \left(\left\lfloor \frac{T}{\tau_i} \right\rfloor + 1 \right) (c_i + d) + \left(\left\lfloor \frac{T}{\tau_s} \right\rfloor + 1 \right) (c_s + d), \right. \\ \left. T d + \sum_{i \in \mathcal{N}} \left(\left\lfloor \frac{T}{\tau_i} \right\rfloor + 1 \right) c_i + \left(\left\lfloor \frac{T}{\tau_s} \right\rfloor + 1 \right) c_s \right\} ,$$

and

$$f_{min} \geq \hat{f}_{min} = \left\lfloor \frac{T}{\max_i(\tau_i)} \right\rfloor d + \sum_{i \in \mathcal{N}} \left\lfloor \frac{T}{\tau_i} \right\rfloor c_i,$$

where $[x]$ is the integer part of x . Thus C is bounded. \square

3 Different discretization approaches

When modeling a discretization of $G(u), u \in [0, \infty)$ with $n \leq T$ points of support, the following questions arise:

1. How many points of support should we use?
2. Which points of support $\kappa_n \subset \kappa_T$ should we choose?
3. How should we place the probability mass?

Answering questions 2 and 3 simultaneously may lead to optimization problems, that are as difficult to solve as the original optimization problem. In section 4 we try to answer question 1 by simulation.

We will describe four different approaches and describe how existing methods can be adapted to our situation. The methods are first presented without any restrictions and then with our restrictions.

3.1 Method minimizing the Sup-distance

Theorem 1 bounds the error measure (20) by a constant times the sup-distance between G_n and G_T . Here we discuss a discretization that minimizes the sup-distance between G_n and G . The reason is that when T tends to infinity, G_T tends to G uniformly in order to be a sequence of discrete approximations of G . The triangle and inverse triangle inequalities gives

$$\begin{aligned} \sup_u |G_n(u) - G(u)| &\leq \sup_u |G_n(u) - G_T(u)| + \sup_u |G_T(u) - G(u)| \\ &= \sup_u |G_n(u) - G_T(u)| + \varepsilon, \end{aligned}$$

and

$$\begin{aligned} \sup_u |G_n(u) - G(u)| &\geq \sup_u |G_n(u) - G_T(u)| - \sup_u |G_T(u) - G(u)| \\ &= \sup_u |G_n(u) - G(u)_T| - \varepsilon, \end{aligned}$$

which gives

$$\sup_u |G_n(u) - G_T(u)| - \varepsilon \leq \sup_u |G_n(u) - G(u)| \leq \sup_u |G_n(u) - G_T(u)| + \varepsilon,$$

where $\varepsilon = \sup_u |G_T(u) - G(u)|$ is small when T is large.

Assume that we do not have a restriction that the points of support have to be in the set κ_T . Then we can minimize $\sup_u |G(u) - G_n(u)|$ where G_n is a discretization with n points of support with masses p_i at u_i , $i = 1 \dots, n$ by choosing point of support i as

$$u_i = G^{-1}\left(\frac{2i-1}{2n}\right) \quad \text{and} \quad p_i = \frac{1}{n}. \quad (22)$$

This can be realized by looking at Figure 1, where a,b,c,d,e and f indicate the greatest difference between G and G_n in intervals 1,2,3 and 4. The overall sup-distance is equal to the maximum distance of a,b,c,d,e or f and is minimized if all distances a,b,c,d,e and f are equal. This gives a sup-distance of $(2n)^{-1}$.

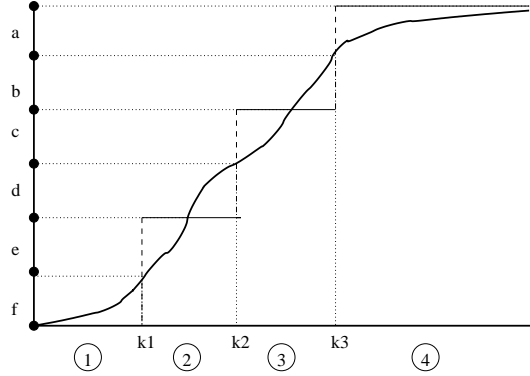


Figure 1: A fictive distribution G and G_3 where a,b,c,d and f are the maximum sup-distance between the functions in intervals 1,2,3 and 4.

With our restrictions on the points of support we get a greater sup-distance. Assume we fix the n points of support in κ_T and form discretization (17). Let $\mathbf{p} = (p_1, \dots, p_n)$ be the probability vector and $\hat{\mathbf{p}}$ the solution

to

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmin}} \{ \sup_u |G(u) - G_n(u)| \}. \quad (23)$$

In most cases $\hat{\mathbf{p}}$ is not unique. We then choose the solution that minimizes the sup-distance in every interval $[(0, k_1), (k_1, k_2), \dots, (k_n, \infty)]$. The overall sup-distance is the maximum of the sup-distance in those intervals so the solution also minimizes the overall sup-distance (23). The solution is as follows

$$\begin{aligned} p_1 &= \frac{G(k_1) + G(k_2)}{2}, \\ p_i &= \frac{G(k_i) + G(k_{i+1})}{2} - \frac{G(k_{i-1}) + G(k_i)}{2} = \frac{G(k_{i+1}) - G(k_{i-1})}{2}, \\ p_n &= 1 - \frac{G(k_n) + G(k_{n-1})}{2}, \end{aligned}$$

where $i = 1, \dots, n-1$. In the intervals $(0, k_1)$ and (k_n, ∞) the sup-distance will be $G(k_1)$ and $1 - G(k_n)$ respectively regardless of $\hat{\mathbf{p}}$. In interval (k_i, k_{i+1}) the minimum sup-distance is $(G(k_{i+1}) - G(k_i))/2$ and it is achieved since

$$\begin{aligned} G_n(k_i) &= \sum_{j=1}^i p_j = \frac{G(k_1) + G(k_2)}{2} + \sum_{j=2}^i \frac{G(k_{j+1}) - G(k_{j-1})}{2} \\ &= \frac{G(k_i) + G(k_{i+1})}{2}, \end{aligned}$$

and

$$\sup_{u \in (k_i, k_{i+1})} |G_n(k_i) - G(u)| = \frac{G(k_{i+1}) - G(k_i)}{2}.$$

If we have the liberty to choose both p_i and which n points of support $k_i \in \kappa_T$ we want to use we get the following problem. Let $(p_i, k_i), i = 1, \dots, n$ be the discretization with p_i the mass in node k_i . Let $\hat{\mathbf{p}}, \hat{\kappa}_n$ be the solution to

$$(\hat{\mathbf{p}}, \hat{\kappa}_n) = \underset{\mathbf{p}, k_i \in \kappa_T}{\operatorname{argmin}} \{ \sup |G(u) - G_n(u)| \}. \quad (24)$$

As above the solution is not unique. Solving this problem is the same as solving problem (23) for all $\binom{T}{n}$ choices of points of support. (Many combinations are unlikely to be best and can be omitted.) In general a good set of points of support are points that lie dense where G increases much.

Instead of the above optimization we suggest to choose points of support near the points in (22). This can be achieved by choosing the n points of support as follows.

1. Choose the first point of support as

$$k_1 = \operatorname{argmin}_{u \in \kappa_T} \sup |G(u) - \frac{1}{2n}|.$$

2. Choose the i 'th point of support as

$$k_i = \operatorname{argmin}_{u \in \kappa_T \cap u > k_{i-1}} \sup |G(u) - \frac{2(1 - G(k_{i-1}))}{2(n - i + 1) + 1} + G(k_{i-1})|. \quad (25)$$

Here we try to spread $G(k_1), \dots, G(k_n)$ uniformly in the interval $[0, 1]$ but since k_i has to be from set κ_T this is not always possible. Equation (25) considers the remaining part of the interval $[0, 1]$, namely $[G(k_{i-1}), 1]$ and spreads the points uniformly in it. Here we start from the beginning of the interval $[0, 1]$, but it would be equally good to start from the end. Note that if there is no restriction of possible points, this choice coincides with (22). Now probabilities can be chosen by (23).

3.2 Bracket means method

This method consists of dividing G into n intervals $\{[t_0, t_1], \dots, [t_{n-1}, t_n]\}$ where $t_0 = 0$, $t_n = \infty$. We now make a discretization by putting the mass

$$p_i = \int_{t_{i-1}}^{t_i} dG(t),$$

at the point

$$u_i = \frac{\int_{t_{i-1}}^{t_i} t dG(t)}{\int_{t_{i-1}}^{t_i} dG(t)}.$$

The intervals can be chosen in many ways. One common choice is that all points of support have equal probability mass. A similar method is the bracket median method, in which one uses the median in each subinterval instead of the mean, see Smith (13).

In the case where the points of support have to be in the set κ_T , it is not always possible to find subintervals so the mean in each subinterval corresponds with a node in κ_T . A simple example that shows this is this: One interval $[0, \infty)$ and mean in distribution does not match node in κ_T .

We present two ways to find approximations that almost satisfy the bracket mean condition. For the first approach we decide which points of support we should use. In the second approach we use a probability vector that helps to choose points. Both methods are expressed as optimization problems that are rewritten in a standard form in appendix A.

Approach 1: Let μ_1 be a metric measuring the distance between point sets. Fix the n nodes $\kappa_n \subset \kappa_T$. Let $\tilde{\kappa}_n = \{\tilde{u}_1, \dots, \tilde{u}_n\}$ be a set of points of support, that is not a subset of κ_T . The solution $\hat{\mathbf{p}} = \{p_1, \dots, p_n\}$ is obtained for the optimization problem

$$\hat{\mathbf{p}} = \underset{\mathbf{p} \in \phi}{\operatorname{argmin}} \mu_1(\kappa_n, \tilde{\kappa}_n) \quad (26)$$

where ϕ is the set of points \mathbf{p} such that

$$\begin{aligned} p_i &= \int_{t_{i-1}}^{t_i} dG(t), \quad i = 1, \dots, n, \\ \tilde{u}_i &= \frac{1}{p_i} \int_{t_{i-1}}^{t_i} t dG(t), \quad i = 1, \dots, n, \\ t_i &> t_{i-1}, \quad i = 1, \dots, n, \\ t_0 &= 0, \\ t_n &= \infty. \end{aligned}$$

Solving (26) gives the discretization of the distribution as in (17). A possible metric μ_1 is $\mu_1 = \sum_{i=1}^n w_i (k_i - \tilde{u}_i)^2$ where w_i is a weight.

Approach 2: Let μ_{21} and μ_{22} be metrics with μ_{21} measuring the distance between probability vectors and μ_{22} measuring the distance between point sets. Let $\mathbf{q} = \{q_1, \dots, q_n\}$ be a probability vector with desirable probabilities and let $\tilde{\kappa}_n = \{\tilde{u}_1, \dots, \tilde{u}_n\}$ be a set of points of support, that is not a subset of κ_T . Then a discretization $\mathbf{p} = \{p_1, \dots, p_n\}$ with points of support $\kappa_n = \{k_1, \dots, k_n\}$ is obtained as the solution $(\hat{\mathbf{p}}, \hat{\kappa}_n)$ to the optimization problem

$$(\hat{\mathbf{p}}, \hat{\kappa}_n) = \underset{\mathbf{p} \in \phi, \kappa_n \subset \kappa_T}{\operatorname{argmin}} \mu_{21}(\mathbf{q}, \mathbf{p}) + \mu_{22}(\kappa_n, \tilde{\kappa}_n), \quad (27)$$

where ϕ is the same set as in the first approach. The solution of (27) gives the discretization of the distribution as in (17). Metrics μ_{21} and μ_{22} need to be chosen so that they do not completely dominate each other. If μ_{21} dominates, we get the brackets with the probabilities we desire but the mean in each bracket may be far from a node. If μ_{22} dominates, the mean in each bracket will be near a node but the probabilities will be far from those desired.

3.3 Method minimizing Wasserstein distance

Pflug (11) suggests the following discretization

$$\begin{aligned} p_1 &= \int_0^{\frac{k_1+k_2}{2}} dG(t), \\ p_i &= \int_{\frac{k_{i-1}+k_i}{2}}^{\frac{k_i+k_{i+1}}{2}} dG(t), \quad i = 2, \dots, n-1, \\ p_n &= \int_{\frac{k_{n-1}+k_n}{2}}^{\infty} dG(t), \end{aligned}$$

where $\{k_1, \dots, k_n\} \in \kappa_n$ are the points of support of the given G_n , derived as the discrete distribution minimizing the Wasserstein distance, see Pflug (11). In Pflug (11) an error measure similar to equation (20) is used, for which he presents a bound depending on the Lipschitz continuity of f with respect to u . His result is not directly applicable to our problem since f is not Lipschitz continuous but we will nevertheless use his method to decide probabilities for comparison purposes.

3.4 Moment preserving method

If we have no constraints on the points of support and the first $2n-1$ moments of the distribution G are finite, then it is possible to create a discrete approximation with n points of support that correctly matches $2n-1$ moments. Let

$$M_j = \int_{-\infty}^{\infty} u^j dG(u),$$

be the j 'th moment. The discretization can be obtained by finding u_i and p_i that satisfy

$$\sum_{i=1}^n p_i u_i^j = M_j \quad j = 0, \dots, 2n-1. \quad (28)$$

For a solution see Miller and Rice(10) and Smith (12). It can be shown that if all M_j are finite and from a probability distribution that spans $[a, b]$ then all u_i will be real and lie in the interval $[a, b]$ and all $p_i \geq 0$.

In our problem, we must choose points from the set κ_T . In order to compare the methods we will use the same points as for the method that minimizes the sup-distance.

Another approach is to calculate points of support using (28) and then choose the points of support in κ_T that are closest to the points obtained. For some distributions this method resulted in points very far out in the tail.

We have n fixed points of support $u_i, i = 1, \dots, n$ and we want to find probabilities $p_i, i = 1, \dots, n$ so that the discretization matches as many moments as possible. Following a simple argument of degrees of freedom we see that the maximum number of moments we can approximate is $n - 1$, (if M_0 is not counted as a moment). Thus we look for a solution p_1, \dots, p_n to

$$\sum_{i=1}^n p_i u_i^j = M_j \quad j = 0, \dots, n - 1. \quad (29)$$

There is no guarantee that there exists a solution with $p_i \geq 0, i = 1, \dots, n$. If some $p_i < 0$ we suggest removing one constraint. The least important constraint is usually the highest order moment constraint

$$\sum_{i=1}^n p_i u_i^{n-1} = M_{n-1}.$$

Removing this we obtain the equations (29) with $j = 0, \dots, n - 2$. If the solution to these satisfies $p_i \geq 0, i = 1, \dots, n$ we are done. If not, we should remove one more moment constraint and keep removing constraints until a solution is found resulting in $p_i \geq 0, i = 1 \dots n$.

After removing moment constraints the solution to the problem is not necessarily unique. (There are more variables than equations). In order to choose one solution we can use a function z that represents some other desired properties in the discretization and solve

$$\hat{\mathbf{p}} = \operatorname{argmax}_{\mathbf{p} \in \varphi} z(\mathbf{p}),$$

where φ is the set of points $\mathbf{p} = (p_1, \dots, p_n)$ such that

$$\begin{aligned} \sum_{i=1}^n p_i u_i^j &= M_j, \quad j = 0, \dots, n - 1 - m, \\ p_i &\geq 0, \quad i = 1, \dots, n, \end{aligned}$$

where m is the number of removed moment constraints. The function z can ,e.g., be formulated to promote p_i of the same sizes if that is a desired property.

4 Test results

In this section we will present numerical results illustrating the different discretization methods in section 3. We will use a Weibull distribution

to describe the life of the stochastic component. We use the following parameterization,

$$\check{G}(u) = 1 - e^{-(\frac{1}{\theta}u)^\alpha}, \quad t > 0, \quad (\theta > 0, \alpha > 0) \quad (30)$$

where θ is the characteristic life and α is the shape parameter. The maximum number of time nodes, T , was set to 30. The distance between the time nodes was set to one and θ was set to 9. Tests were made with the α parameter being both 1 and 2.

In the following we will make an attempt to establish a rule of thumb for how many points of support are needed. We will also comment on the consequence of using a narrow scenario tree.

4.1 Error measure

We study how the error measure depends on different discretizations methods and different number of points of support. We model the engine with two components, one stochastic and the other deterministic. When determining components to replace, there are four alternatives:

1. Replace the deterministic component.
2. Replace the stochastic component.
3. Replace both components.
4. Do not replace any components.

The optimal replacement alternative was calculated in optimization problem (19) with $n = 1, \dots, 10$ points of support.

The points of support were chosen by equation (25). The probabilities were chosen in four ways, by the method that minimizes the sup-distance, the method that minimizes the Wasserstein distances, the method that preserves the moments, and the bracket method approach 1.

The best discretization possible was a discretization with 30 points of support, one in every node, for which the optimal replacement alternative was calculated according to (18). Finally the difference between the two discretizations, using error measure (20), was calculated. The calculation was performed in AMPL (a modeling language for mathematical programming).

The parameters and remaining lives of the components used are shown in Table 1.

The expected life of the stochastic component τ_s is connected to the shape parameter α in (30). The parameter τ_s was set to 9 when $\alpha = 1$ and 8 when $\alpha = 2$. The age of the stochastic component was changed

Parameter	Alternatives
$\tilde{\tau}_1$	4, 6
τ_1	6, 10
c_1	60, 100, 130
c_s	70, 100, 150
d	70, 100, 150
α	1, 2

Table 1: Parameters changed in the optimization model (2) and Weibull distribution (30).

from new to an age of 9 in four steps, (new, 3, 6, 9). The error measure was calculated for all levels of the parameters, ages of the stochastic component and discretization methods. In Figure 2 the result of the test is presented.

The method that preserves the moments seems worse and the methods that minimize the Wasserstein and Sup-distance seem better. It seems that using two points of support is worse than just using the expected value of the distribution. At three points of support or more the error measure seems rather constant, compared with using one or two points of support, if we disregard the moment method with $\alpha = 1$. Some work has been done on three points of support discrete distributions, cf. Keefer (9).

Further tests showed that the error measure decreases as α increases. With constant θ , the variance of the Weibull distribution (30) decreases with the increasing α parameter. If the variance is great it is harder to describe the distribution with just a few points of support.

4.2 Narrow tree approximation

The greatest approximation is that only the first life of the stochastic component is modeled with several points of support. After the stochastic component is replaced it is modeled with the expected value of the life distribution, which is the same as using one point of support. The reason why this approximation was invented in the first place was that it was natural to think that what happens in the near future has a greater impact on the decision we have to make today, than what is going to happen a long time from now. The optimization model works in a different way. Basically it tries to find a scheme with replacement times that, for the entire service period, minimize the expected cost of maintaining the engine. The approximations sometimes lead to irrational solutions, such as when a new stochastic detail is replaced.

A simple example that illustrates this is the following one. Assume we have one deterministic and one stochastic detail. The engine is at the

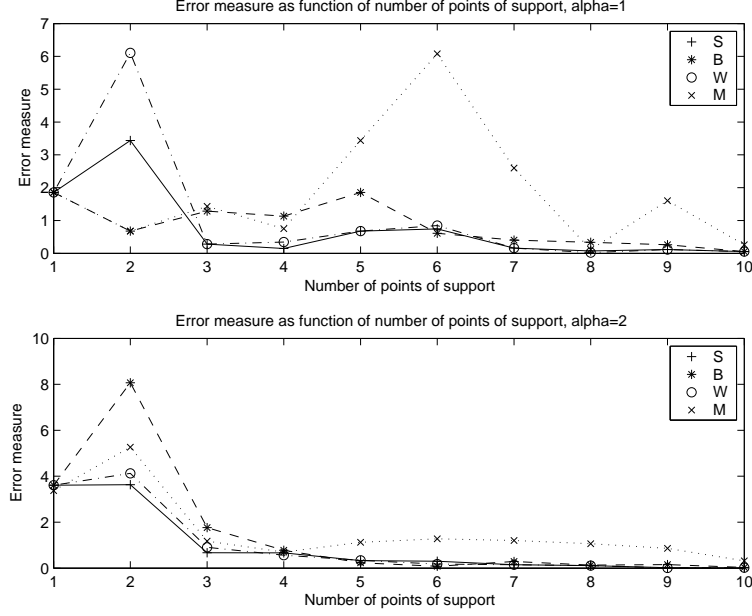


Figure 2: *Y-axis is the mean error measure with parameters as in Table 1. X-axis is the number of points of support, S = method that minimizes the sup-distance, B = Bracket method, W = method that minimize Wasserstein distance, M = Moment preserving method.*

repair bay because the deterministic detail needs replacement. For the sake of simplicity assume that the stochastic component is new, that is $\tilde{G} = G$. We use the model to answer the question if we should replace the stochastic detail. Assume the price for making repairs is $z_t = 20$ and that replacing a component costs $c_1 = c_s = 10$. We model the stochastic component with 2 points of support. It fails in node 3 with probability $p_A = 0.5$ and in node 5 with probability $p_B = 0.5$. The expected value of \tilde{G} corresponds to a distance of 4 nodes. Consequently the stochastic component that replaces the first stochastic component must be replaced at least at every four nodes. Assume that the deterministic component has a life of 4 nodes. The total time we need to maintain the engine is $T = 8$ nodes. Note that there is no restriction that the engine needs to function at node 8. In figure 3 we see three replacement schemes.

In node zero we see the first stage variables, that is the decision we make now. Given that we have to replace the deterministic component we

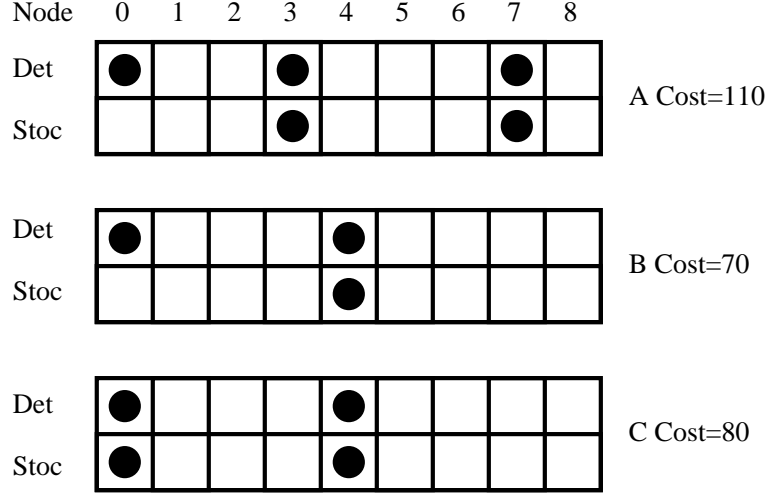


Figure 3: Three replacement schemes in an optimization problem that contains one deterministic and one stochastic component. A dot indicates replacement. The prices of replacing a component is 10 and the cost of bringing the engine to service is 20

can either keep the stochastic component (schemes A and B) or replace it (scheme C). For each scheme a cost q is calculated. The model will suggest a replacement of the stochastic component if the cost $q_C < p_A q_A + p_B q_B$. With our prices this is $80 < 90$. Consequently we replace the stochastic component even if it is assumed to be new.

In the case where $\alpha = 1$ the Weibull distribution (30) becomes an exponential distribution. The exponential distribution has a constant failure or hazard rate, which means that the risk of failure is constant over time. Thus a stochastic component with an exponential distribution never needs replacement. Table 2 contains results on the percent of replacements from the test run with the exponential distribution. It seems that the model very often replaces the component even if replacement is not necessary. The problem is not solved by increasing the number of points of support. When 30 points of support is used the stochastic component is still replaced in 18% of the times. When only one point of support is used there is no replacement of the stochastic component because the replacing component is modeled in the same way. This model problem can probably be corrected if more than the first stochastic life distribution is modeled with several points of support.

PoS	S	B	W	M
1	0	0	0	0
2	0.44	0.11	0.56	0.11
3	0.10	0.31	0.10	0
4	0.15	0.20	0.19	0.06
5	0.11	0	0.11	0.44
6	0.31	0.22	0.31	0.55
7	0.24	0.27	0.24	0.41
8	0.12	0.09	0.13	0.24
9	0.24	0.26	0.24	0.37
10	0.13	0.13	0.13	0.26

Table 2: The percent of times the optimization model replaced the stochastic component even if it was not aging. PoS=Points of Support, S = method that minimizes Sup-distance, B = Bracket method, W = method that minimizes Wasserstein distance, M = Moment preserving method.

5 Summary and further work

An optimization model for replacement of parts in an aircraft engine containing stochastic and deterministic components has been studied. We made a discretization of the distribution of the life of the stochastic component. The structure of the optimization model demands that the points of support of the discretization coincide with the nodes in the model. Four different discretization methods were presented, the bracket mean method, the moment preserving method, a Wasserstein distance minimizing method, and a method that minimizes the Sup-distance. The discretization methods were adapted to the constraints in the optimization model.

An error measure closely connected to the optimization model was defined and a maximum limitation of the error was derived. With the help of the error measure and test runs the different methods were compared. According to the test, the methods that minimize the Wasserstein and Sup-distance were better. The moment preserving method performed worse. The minimum number of points of support suggested is three.

When using the narrow scenario tree many details were replaced even if they did not need replacement. This model problem can probably be corrected if more than the first stochastic life distribution is modeled with several points of support.

Possible future work is to model more lives with several points of support. Furthermore we need to study a model with more stochastic details and how to make discretizations if the lives of the components are correlated.

Acknowledgements

The author wants to thank Dragi Anevski at Göteborg University and Jacques de Maré at Chalmers for supervision. This work was financed by NFFP.

Appendix

A: Bracket method. The optimization problems in section 3.2 is here rewritten in a less explicit form. Let

$$H(\mathbf{t}) = \begin{pmatrix} H_1(t_1, t_2) \\ \vdots \\ H_i(t_{i-1}, t_i) \\ \vdots \\ H_n(t_{n-1}, t_n) \end{pmatrix} = \begin{pmatrix} \int_{t_0}^{t_1} dG(u) \\ \vdots \\ \int_{t_{i-1}}^{t_i} dG(u) \\ \vdots \\ \int_{t_{n-1}}^{t_n} dG(u) \end{pmatrix}, \quad (31)$$

$$B(\mathbf{t}) = \begin{pmatrix} B_1(t_1, t_2) \\ \vdots \\ B_i(t_{i-1}, t_i) \\ \vdots \\ B_n(t_{n-1}, t_n) \end{pmatrix} = \begin{pmatrix} \frac{\int_{t_0}^{t_1} t dG(u)}{H_1(t_1, t_2)} \\ \vdots \\ \frac{\int_{t_{i-1}}^{t_i} t dG(u)}{H_i(t_{i-1}, t_i)} \\ \vdots \\ \frac{\int_{t_{n-1}}^{t_n} t dG(u)}{H_n(t_{n-1}, t_n)} \end{pmatrix}, \quad (32)$$

where $\mathbf{t} = (t_1, \dots, t_n)$. Problem (26) can now be formulated

$$\bar{\mathbf{t}} = \operatorname{argmin}_{\mathbf{t} \in \Gamma} \mu_1(\kappa_n, B(\mathbf{t})), \quad (33)$$

and the problem (27) and be formulated

$$(\bar{\mathbf{t}}, \bar{\kappa}_n) = \operatorname{argmin}_{\kappa_n \subset \kappa_T, \mathbf{t} \in \Gamma} \mu_{21}(\mathbf{p}, H(\mathbf{t})) + \mu_{22}(\kappa_n, B(\mathbf{t})), \quad (34)$$

where Γ is the set of points \mathbf{t} such that

$$\begin{aligned} t_i &\geq t_{i-1} + \epsilon, \quad i = 1, \dots, n, \\ t_0 &= 0, \\ t_n &= G^{-1}(p_x), \end{aligned}$$

where ϵ is a positive small number and p_x is a probability near 1. If nice metrics are chosen the objective function will be continuous. The

constraints form a compact set in \mathbf{t} . Problem (33) is solved by standard methods. Problem (34) can be solved for every fix choice of $\kappa_n \subset \kappa_T$. We can make $\binom{T}{n}$ different choices of κ_n and need to check which one minimizes the objective function in (34). For large T this is computationally demanding.

References

- [1] F Altenstedt. *Aspects on asset liability management via stochastic programming*. PhD thesis, Chalmers University of Technology, 2003.
- [2] N. Andréasson. Optimization of opportunistic replacement activities in deterministic and stochastic multi-component systems. Licentiate thesis, Chalmers University of Technology, 2004.
- [3] Jitka Dupačová, Giorgio Consigli, and Stein W. Wallace. Scenarios for multistage stochastic programs. *Ann. Oper. Res.*, 100:25–53 (2001), 2000. Research in stochastic programming (Vancouver, BC, 1998).
- [4] Y. Epstein, S. Wilamowsky. Opportunistic replacement in a deterministic environment. *Comput. and Opns. Res.*, 12:311–322, 1985.
- [5] B. Dickman et al. Modeling deterministic opportunistic replacement as an integer programming problem. *Amer. J.Math. and Mgmt.*, 11, 1991.
- [6] Karl Frauendorfer and Michael Schürle. Term structure models in multistage stochastic programming: estimation and approximation. *Ann. Oper. Res.*, 100:189–209 (2001), 2000. Research in stochastic programming (Vancouver, BC, 1998).
- [7] W. Høyland, K. Wallace. Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307, 2001.
- [8] Peter Kall and Stein W. Wallace. *Stochastic programming*. Wiley-Interscience Series in Systems and Optimization. John Wiley & Sons Ltd., Chichester, 1994.
- [9] D.L. Keefer. Certain equivalents for three-point discrete-distribution approximations. *Management Science*, 40:760–773, 1994.
- [10] T.R. Miller, A.C. Rice. Discrete approximations to probability distributions. *Management Science*, 29:352–362, 1983.

- [11] G. Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Math. Program.*, 89(2, Ser. B):251–271, 2001. Mathematical programming and finance.
- [12] J.E. Smith. *Moment method for decision analysis*. PhD thesis, Stanford University, 1990.
- [13] J.E. Smith. Moment methods for decision analysis. *Management Science*, 39:340–358, 1993.
- [14] H. Wang. A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research*, 139:469–489, 2002.