Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

Lecture Notes: Reliability Engineering

Quality Management 444

Department of Industrial Engineering

Dr JL Jooste*

Last Revision: July 24, 2023

*Contact details: Faculty of Engineering, Industrial Engineering Building, Room 5019, Joubert Street, Stellenbosch, 7600; Tel: +27 21 808 4234; Email: wyhan@sun.ac.za

forward together
sonke siya phambili
saam vorentoe

Contents

Chapter 1

An Introduction to Quality Management and Reliability Engineering

This chapter serves as an introduction about the concepts of quality management and reliability engineering and the synergy between these fields. The chapter contains content predominantly from O'Connor and Kleyner (2012, 1-2, 4-9, 13-18, 421-424) which is a reproduced under Stellenbosch University's blanket copyright licence with the Dramatic, Artistic and Literary Rights Organisation (DALRO) for Lecture Notes: Reliability Engineering in 2023.

1.1 Quality and Reliability Concepts

Based on the type of consumer, **products** can be classified as consumer goods and industrial products. The former is for consumers (i.e. cars, mobile phones) and the latter for other businesses which are often referred to as business-to-business products or capital goods (i.e. engineering machines). A product can be a single component or a complex system consisting of many components. For the latter case, a product can be decomposed into a hierarchy structure typically consisting of components, assemblies, subsystems and systems (Jiang, 2015, 3).

The concept of **product life cycle** is different for manufacturers and consumers. For the manufacturer, the product life cycle consists of the phases of a product's life, from conception, through design and manufacturing to after sales services and disposal. For the consumer, the product life cycle consists of the time from purchasing the product to its discarding when it reaches its end of useful life or when its being replaced earlier due to technological obsolescence or due to the product not being of use any more (Jiang 2015, 4; Murthy et al. 2009) .

Engineered products need to be reliable. The average consumer is acutely aware of the problem of less than perfect reliability in domestic products such as TV sets, vehicles and mobile phones. Further, large asset owning organisations such as airlines and public utilities are aware of the costs of unreliability. Manufacturers often suffer high costs of failure under warranty. Argument and misunderstanding begin when we try to quantify reliability values, or try to assign financial or other cost or benefit values to levels of reliability.

The simplest, purely producer-oriented or inspectors' view of **quality** is that in which **a product is assessed for conformance against a specification or set of attributes**, and when passed is delivered to the customer[1] The customer accepting the product, accepts that it might fail at some future time. This simple approach is often associated with a warranty, or the customer may have some protection in law, so that he may claim back for failures occurring within a stated or reasonable time. However, this approach provides no measure of quality over a period of time, particularly outside a warranty period. Even within a warranty period, the customer usually has no grounds for further action if the product fails once, twice or several times, provided that the manufacturer repairs the product as promised each time. If it fails often, the manufacturer will suffer high warranty costs, and the customers will suffer inconvenience. Outside the warranty period, only the customer suffers and the manufacturer will also probably incur a loss of reputation, possibly affecting future business.

We therefore need a time-based concept of quality. The inspectors' concept is not time-dependent. The product either passes a given test or it fails. On the other hand, reliability is usually concerned with failures in the time domain. This distinction marks the difference between traditional quality control and reliability engineering. Whether failures occur or not, and the times to failure occurrence, can seldom be forecast accurately. Reliability is therefore an aspect of engineering uncertainty. Whether an item will work for a particular period can be expressed as a probability. This results in the usual engineering **definition of reliability** as:

> *The probability that an item will perform a required function without failure under stated conditions for a stated period of time.*

Reliability can also be expressed as the number of failures over a period.

Durability is a particular aspect of reliability, related to the ability of an item to withstand the effects of time (or of distance travelled, operating cycles, etc.) dependent mechanisms such as fatigue, wear, corrosion, electrical parameter change, and so on. Durability is usually expressed as a minimum time before the occurrence of wear-out failures. In repairable systems it often characterizes the ability of the product to function after repairs.

The **objectives of reliability engineering**, in the order of priority, are:

1. To apply engineering knowledge and specialist techniques to prevent or to reduce the likelihood or frequency of failures.
2. To identify and correct the causes of failures that do occur, despite the efforts to prevent them.
3. To determine ways of coping with failures that do occur, if their causes have not been corrected.
4. To apply methods for estimating the likely reliability of new designs, and for analysing reliability data.

The reason for the priority emphasis is that it is by far the most effective way of working, in terms of minimising costs and generating reliable products.

[1] This is a simplistic perspective of quality for illustrating the essential difference between quality and reliability. More elaborate definitions for quality, quality control, quality assurance, quality management and total quality management (TQM) exists. It should also be recognised that the boundary between quality and reliability is more often than not blurred, which will later become apparent in section 1.5.

The primary skills that are required, therefore, are the ability to understand and anticipate the possible causes of failures, and knowledge of how to prevent them. It is also necessary to have knowledge of the methods that can be used for analysing designs and data. The primary skills are nothing more than good engineering knowledge and experience, so reliability engineering is first and foremost the application of good engineering, in the widest sense, during design, development, manufacture and service.

Mathematical and statistical methods can be used for quantifying reliability (prediction, measurement) and for analysing reliability data. However, because of the high levels of uncertainty involved these can seldom be applied with the kind of precision and credibility that engineers are accustomed to when dealing with most other problems. In practice the uncertainty is often in orders of magnitude. Therefore the role of mathematical and statistical methods in reliability engineering is limited, and appreciation of the uncertainty is important in order to minimise the chances of performing inappropriate analysis and of generating misleading results. Mathematical and statistical methods can make valuable contributions in appropriate circumstances, but practical engineering must take precedence in determining the causes of problems and their solutions. Unfortunately not all reliability training, literature and practice reflect this reality.

Above all of these aspects, though, is the management of the reliability engineering effort. Since reliability (and very often also safety) is such a critical parameter of most modern engineering products, and since failures are generated primarily by the people involved (designers, test engineers, manufacturing, suppliers, maintainers, users), it can be maximised only by an integrated effort that encompasses training, teamwork, discipline, and application of the most appropriate methods. Reliability engineering "specialists" cannot make this happen. They can provide support, training and tools, but only managers can organise, motivate, lead and provide the resources. Reliability engineering is, ultimately, effective management of engineering.

1.2 The History of Reliability Engineering

Reliability engineering, as a separate engineering discipline, originated in the United States during the 1950s. The increasing complexity of military electronic systems was generating failure rates which resulted in greatly reduced availability and increased costs. The gathering pace of electronic device technology meant that the developers of new military systems were making increasing use of large numbers of new component types, involving new manufacturing processes, with the inevitable consequences of low reliability. The users of such equipment were also finding that the problems of diagnosing and repairing the new complex equipment were seriously affecting its availability for use, and the costs of spares, training and other logistics support were becoming excessive.

Further, the revolution in electronic device technology continued, led by integrated micro circuitry. Increased emphasis was now placed on improving the quality of devices fitted to production equipments. Screening techniques, in which all production devices are subjected to elevated thermal, electrical and other stresses, were introduced in place of the traditional sampling techniques. With component populations on even single printed circuit boards becoming so large, sampling no longer provided sufficient protection against the production of defective equipment. These techniques were formalised in military standards covering the full range of electronic components. Specifications and test systems for electronic components, based upon the US Military Standards, were developed in the United Kingdom and in continental Europe, and internationally through the International Electro-technical Commission (IEC).

Improved quality standards in the electronic components industry resulted in dramatic improvements in the reliability of commercial components. As a result, during the 1980s the US Military began switching from military grade electronic components to "commercial off the shelf" (COTS) parts in order to reduce costs, and this approach has spread to other applications. Engineering reliability effort in the United States developed quickly, and the reliability programme concepts were adopted by NASA and many other major suppliers and purchasers of high technology equipment.

The concept of life cycle costs (LCC), or whole life costs, was introduced. In the United Kingdom, Defence Standard 00-40, The Management of Reliability and Maintainability was issued in 1981. In the 1990s the series of European Reliability/Dependability standards began to be developed, and became integrated into the International Standards Organization (ISO).

Starting in the early 1980s, the reliability of new Japanese industrial and commercial products took Western competitors by surprise. Products such as automobiles, electronic components and systems, and machine tools achieved levels of reliability far in excess of previous experience. These products were also less expensive and often boasted superior features and performance. The "Japanese quality revolution" had been driven by the lessons taught by American teachers brought in to help Japan's post-war recovery. The two that stand out were J.R. Juran and W. Edwards Deming, who taught the principles of **total quality management** (TQM) and continuous improvement. Japanese pioneers, particularly K. Ishikawa, also contributed. These ideas were all firmly rooted in the teaching of the American writer on management, Peter Drucker, that people work most effectively when they are given the knowledge and authority to identify and implement improvements, rather than being expected to work to methods dictated by management.

The Western approach had been based primarily on formal procedures for design analysis and reliability demonstration testing, whereas the Japanese concentrated on manufacturing quality. Nowadays most customers for systems such as military, telecommunications, transport, power generation and distribution, and so on, rely upon contractual motivation, such as warranties and service support, rather than on imposition of standards that dictate exactly how reliability activities should be performed.

Another aspect of reliability thinking that has developed is the application of statistical methods, primarily to the analysis of failure data and to predictions of reliability and safety of systems. Since reliability can be expressed as a probability, and is affected by variation, in principle these methods are applicable. They form the basis of most teaching and literature on the subject. However, variation in engineering is usually of such an uncertain nature that refined mathematical and quantitative techniques can be inappropriate and misleading.

1.3 Why Do Engineering Products Fail?

There are many reasons why a product might fail. Knowing, as far as is practicable, the potential causes of failures is fundamental to preventing them. It is rarely practicable to anticipate all of the causes, so it is also necessary to take account of the uncertainty involved. The reliability engineering effort, during design, development and in manufacture and service should address all of the anticipated and possibly unanticipated causes of failure, to ensure that their occurrence is prevented or minimised. The main reasons why failures occur are:

1. The design might be inherently incapable. It might be too weak, consume too much power, suffer resonance at the wrong frequency, and so on. The list of possible reasons is endless, and every design problem presents the potential for errors, omissions, and oversights. The more complex the design or difficult the problems to be overcome, the greater is this potential.
2. The item might be overstressed in some way. If the stress applied exceeds the strength then failure

will occur. An electronic component will fail if the applied electrical stress (voltage, current) exceeds the ability to withstand it, and a mechanical strut will buckle if the compression stress applied exceeds the buckling strength. Overstress failures such as these do happen, but fortunately not very often, since designers provide margins of safety. Electronic component specifications state the maximum rated conditions of application, and circuit designers take care that these rated values are not exceeded in service. In most cases they will in fact do what they can to ensure that the in-service worst case stresses remain below the rated stress values: this is called “de-rating”. Mechanical designers work in the same way: they know the properties of the materials being used (e.g. ultimate tensile strength) and they ensure that there is an adequate margin between the strength of the component and the maximum applied stress. However, it might not be possible to provide protection against every possible stress application.

3. Failures might be caused by variation. In the situations described above the values of strength and load are fixed and known. If the known strength always exceeds the known load, as shown in Figure 1.1, then failure will not occur. However, in most cases, there will be some uncertainty about both. The actual strength values of any population of components will vary: there will be some that are relatively strong, others that are relatively weak, but most will be of nearly average strength. Also, the loads applied will be variable. Figure 1.2 shows this type of situation. As before, failure will not occur so long as the applied load does not exceed the strength. However, if there is an overlap between the distributions of load and strength, and a load value in the high tail of the load distribution is applied to an item in the weak tail of the strength distribution so that there is overlap or interference between the distributions (Figure 1.3), then failure will occur.

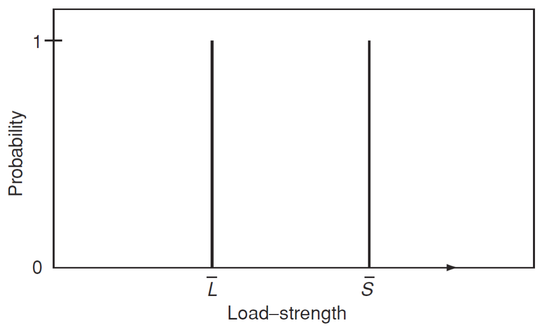


Figure 1.1: Load-strength discrete values

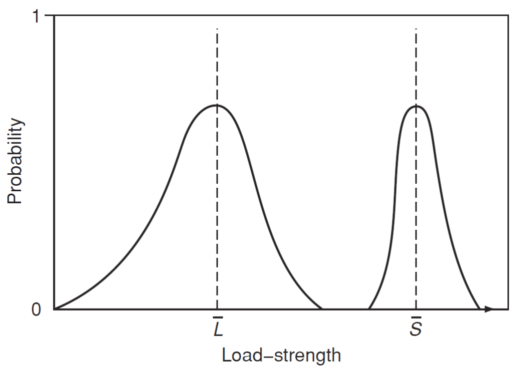


Figure 1.2: Load-strength distributed values

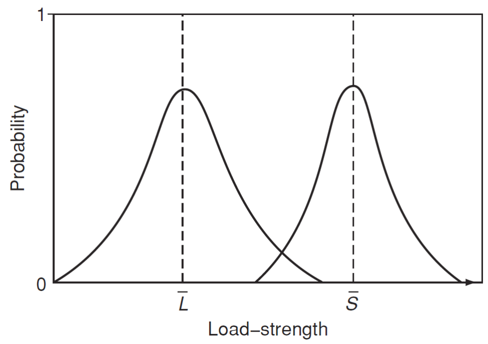


Figure 1.3: Load-strength interfering distributions

4. Failures can be caused by wear-out. We will use this term to include any mechanism or process that causes an item that is sufficiently strong at the start of its life to become weaker with age. Well-known examples of such processes are material fatigue, wear between surfaces in moving contact, corrosion, insulation deterioration, and the wear-out mechanisms of light bulbs and fluorescent tubes. Figure 1.4 illustrates this kind of situation. Initially the strength is adequate to withstand the applied loads, but as weakening occurs over time the strength decreases. In every case the average value falls and the spread of the strength distribution widens. This is a major reason why it is so difficult to provide accurate predictions of the lives of such items.

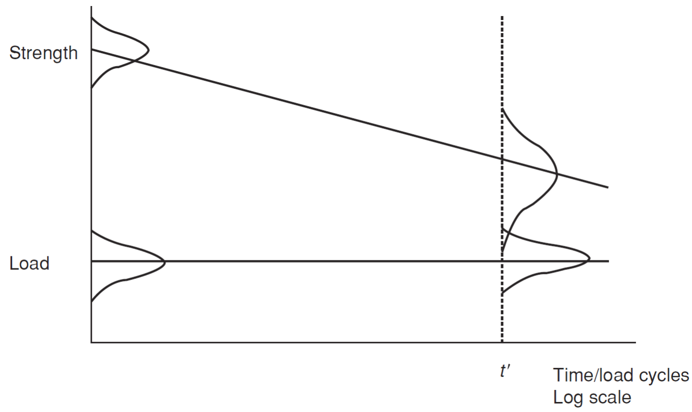


Figure 1.4: Time-dependent load and strength variation

5. Failures can be caused by other time-dependent mechanisms. Battery run-down, creep caused by simultaneous high temperature and tensile stress, as in turbine discs and fine solder joints, and progressive drift of electronic component parameter values are examples of such mechanisms.

6. Failures can be caused by sneaks. A sneak is a condition in which the system does not work properly even though every part does. For example, an electronic system might be designed in such a way that under certain conditions incorrect operation occurs. The fatal fire in the Apollo spacecraft crew capsule was caused in this way: the circuit design ensured that an electrical short circuit would occur when a particular sequence was performed by the crew. Sneaks can also occur in software designs.

7. Failures can be caused by errors, such as incorrect specifications, designs or software coding, by faulty assembly or test, by inadequate or incorrect maintenance, or by incorrect use. The actual failure mechanisms that result might include most of the list above.

8. There are many other potential causes of failure. Gears might be noisy, oil seals might leak, display screens might flicker, operating instructions might be wrong or ambiguous, electronic systems might suffer from electromagnetic interference, and so on.

Failures have many different causes and effects, and there are also different perceptions of what kinds of events might be classified as failures. The burning O-ring seals on the Space Shuttle booster rockets were not considered as failures, until the ill-fated launch of Challenger. We also know that all failures, in principle and almost always in practice, can be prevented.

1.4 Probabilistic Reliability

The concept of reliability as a probability means that any attempt to quantify it must involve the use of statistical methods. An understanding of statistics as applicable to reliability engineering is therefore a necessary basis for progress, except for the special cases when reliability is perfect (we know the item will never fail) or it is zero (the item will never work). In engineering we try to ensure 100% reliability, but our experience tells us that we do not always succeed. Therefore reliability statistics are usually concerned with probability values which are very high (or very low: the probability that a failure does occur, which is 1-reliability). Quantifying such numbers brings increased uncertainty, since we need correspondingly more information. Other sources of uncertainty are introduced because reliability is often about people who make and people who use the product, and because of the widely varying environments in which typical products might operate.

Further uncertainty, often of a subjective nature, is introduced when engineers begin to discuss failures. Should a failure be counted if it was due to an error that is hoped will not be repeated? If design action is taken to reduce the risk of one type of failure, how can we quantify our trust in the designer's success? Was the machine under test typical of the population of machines?

Reliability is quantified in other ways. We can specify a reliability as the mean number of failures in a given time (failure rate), or as the *mean time between failures* (MTBF) for items which are repaired and returned to use, or as the *mean time to failure* (MTTF) for items which are not repaired, or as the proportion of the total population of items failing during the mission life.

The application and interpretation of statistics to deal with the effects of variability on reliability are less straightforward than in, say, public opinion polls or measurement of human variations such as IQ or height. In these applications, most interest is centred around the behaviour of the larger part of the population or sample, variation is not very large and data are plentiful. In reliability we are concerned with the behaviour in the extreme tails of distributions and possibly unlikely combinations of load and strength, where variability is often hard to quantify and data are expensive.

Further difficulties arise in the application of statistical theory to reliability engineering, owing to the fact that variation is often a function of time or of time-related factors such as operating cycles, diurnal or seasonal cycles, maintenance periods, and so on. Engineering, unlike most fields of knowledge, is primarily concerned with change, hopefully, but not always, for the better. Therefore the reliability data from any past situation cannot be used to make credible forecasts of the future behaviour, without taking into account non-statistical factors such as design changes, maintainer training, and even imponderables such as unforeseeable production or service problems. The statistician working in reliability engineering needs to be aware of these realities.

It must always be remembered that quality and reliability data contain many sources of uncertainty and variability which cannot be rigorously quantified. It is also important to appreciate that failures and their causes are by no means always clear-cut and unambiguous. They are often open to interpretation and argument. They also differ in terms of importance (cost, safety, other effects). Therefore we must be careful not to apply only conventional scientific, deterministic thinking to the interpretation of failures. For example, a mere count of total reported failures of a product is seldom useful or revealing. It tells us nothing about causes or consequences, and therefore nothing about how to improve the situation. This contrasts with a statement of a physical attribute such as weight or power consumption, which is usually unambiguous and complete. Nevertheless, it is necessary to derive values for decision-making, so the mathematics are essential. The important point is that the reliability engineer or manager is not, like an insurance actuary, a powerless observer of his statistics. Statistical derivations of reliability are not a guarantee of results, and these results can be significantly affected by actions taken by quality and reliability engineers and managers.

1.5 Reliability and Quality Management Programme Activities

What, then, are the actions that managers and engineers can take to influence reliability? One obvious activity already mentioned is **quality assurance** (QA), the whole range of functions designed to ensure that delivered products are compliant with the design. For many products, QA is sufficient to ensure high reliability, and we would not expect a company mass-producing simple die castings for non-critical applications to employ reliability staff. In such cases the designs are simple and well proven, the environments in which the products will operate are well understood and the very occasional failure has no significant financial or operational effect. QA, together with craftsmanship, can provide adequate assurance for simple products or when the risks are known to be very low. Risks are low when safety margins can be made very large, as in most structural engineering. Reliability engineering disciplines may justifiably be absent in many types of product development and manufacture. QA disciplines are, however, essential elements of any integrated reliability programme.

The reliability effort should always be treated as an integral part of the product development and not as a parallel activity unresponsive to the rest of the development programme. This is the major justification for placing responsibility for reliability with the project manager. Whilst specialist reliability services and support can be provided from a central department in a matrix management structure, the responsibility for reliability achievement must not be taken away from the project manager, who is the only person who can ensure that the right balance is struck in allocating resources and time between the various competing aspects of product development.

The elements of a comprehensive and integrated quality and reliability programme are shown, related to the overall development, production and in-service programme, in Figure 1.5. These show the continuous feedback of information, so that design iteration can be most effective. It can be seen that Most of the design for reliability (DfR) tools and activities feature in the Figure 1.5 flow.

Since production quality will affect reliability, quality control is an integral part of the programme. Quality control cannot make up for design shortfalls, but poor quality can negate much of the reliability only at reducing production costs and the passing of a final test or inspection. Quality control can be made to contribute most effectively to the reliability effort if:

1. 1 Quality procedures, such as test and inspection criteria, are related to factors which can affect reliability, and not only to form and function. Examples are tolerances, inspection for flaws which can cause weakening, and the need for adequate screening when appropriate.

2. Quality control test and inspection data are integrated with the other reliability data.

3. Quality control personnel are trained to recognize the relevance of their work to reliability, and trained and motivated to contribute.

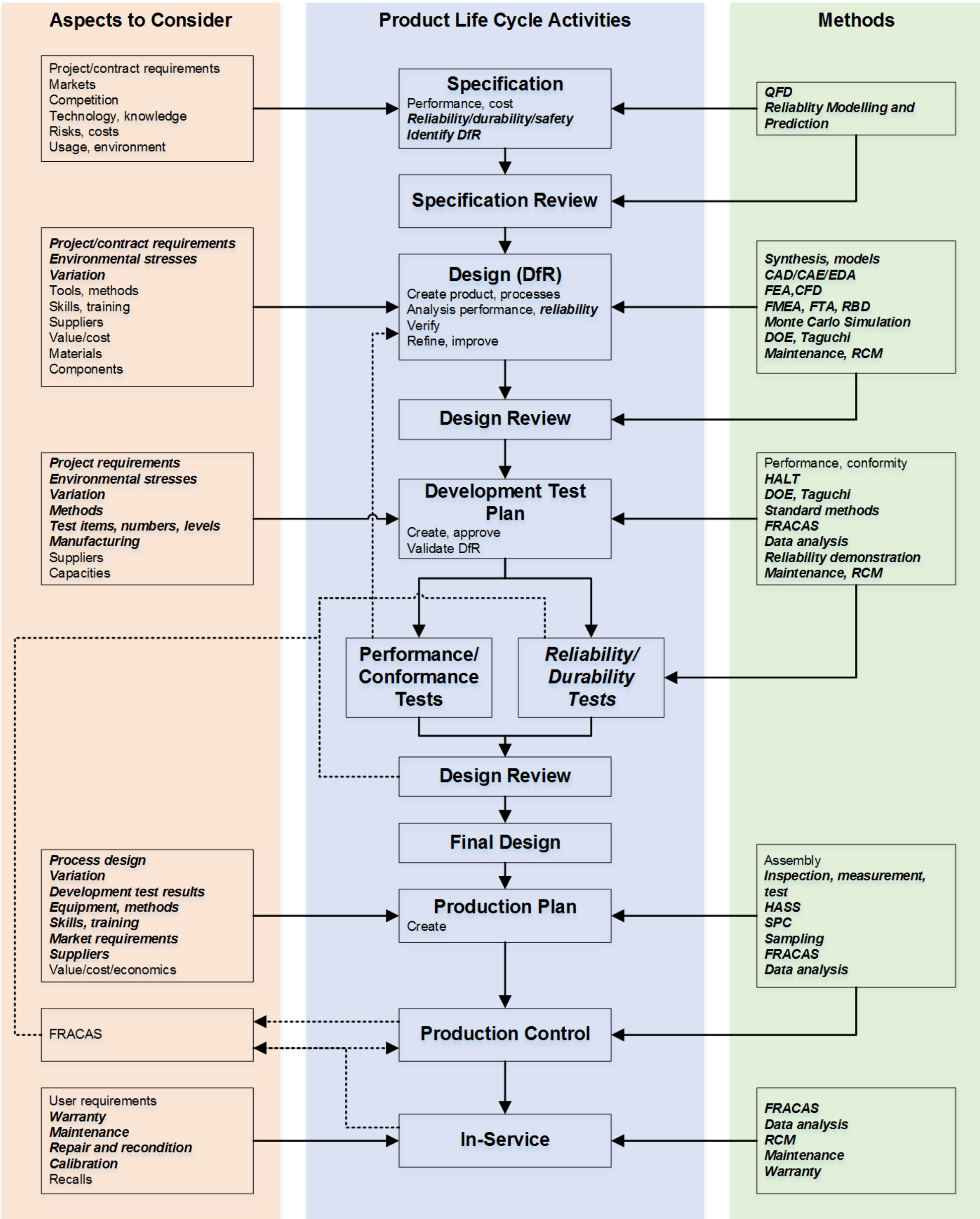


Figure 1.5: Integrated quality and reliability programme flow (items in bold italics indicate quality and reliability specific aspects) (Adapted from O'Connor and Kleyner (2012))

An integrated reliability programme must be disciplined. Whilst creative work such as design is usually most effective when not constrained by too many rules and guidelines, the reliability (and quality) effort must be tightly controlled and supported by mandatory procedures. The disciplines of design analysis, test, reporting, failure analysis and corrective action must be strictly imposed, since any relaxation can result in a reduction of reliability, without any reduction in the cost of the programme. There will always be pressure to relax the severity of design analyses or to classify a failure as non-relevant if doubt exists, but this must be resisted. The most effective way to ensure this is to have the agreed the programme activities written down as mandatory procedures, with defined responsibilities for completing and reporting all tasks, and to check by audit and during programme reviews that they have been carried out.

1.6 Reliability Economics and Management

Obviously the reliability programme activities described can be expensive. Figure 1.6 is a commonly described representation of the theoretical cost-benefit relationship of effort expended on reliability (and production quality) activities. It shows a U-shaped total cost curve with the minimum cost occurring at a reliability level somewhat lower than 100%. This would be the optimum reliability, from the total cost point of view. W.E. Deming presented a different model in his teaching on manufacturing quality. This is shown in Figure 1.7 He argued that, since less than perfect quality is the result of failures, all of which have causes, we should not be tempted to assume that any level of quality is "optimum", but should ask "what is the cost of preventing or correcting the causes, on a case by case basis, compared with the cost of doing nothing?" When each potential or actual cause is analysed in this way, it is usually found that it costs less to correct the causes than to do nothing. Thus total costs continue to reduce as quality is improved. This simple picture was the prime determinant of the post-war quality revolution in Japan, and formed the basis for the philosophy of kaizen (continuous improvement). 100% quality was rarely achieved, but the levels that were achieved exceeded those of most Western competitors, and production costs were reduced.

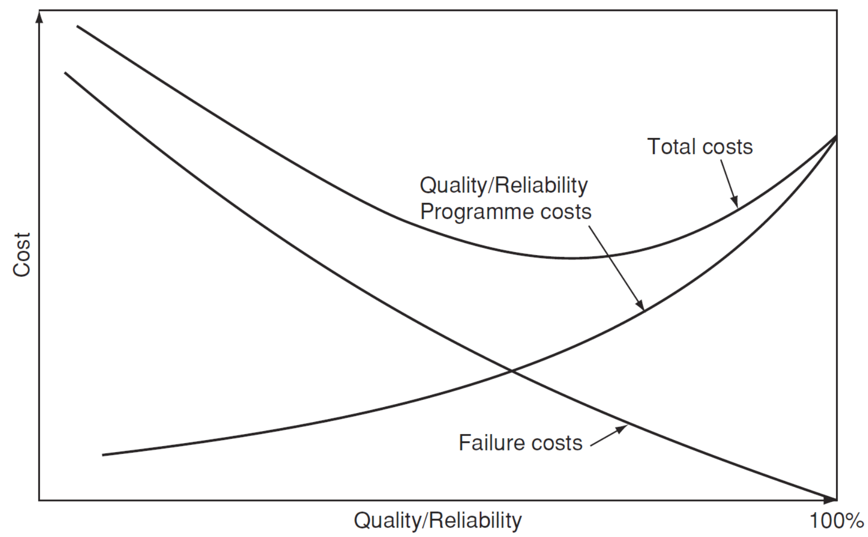


Figure 1.6: Reliability and life cycle costs (traditional view)

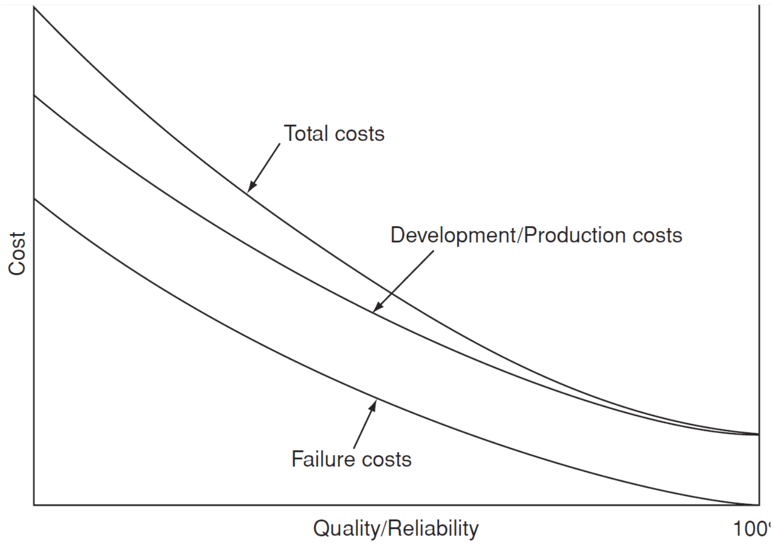


Figure 1.7: Reliability/Quality and life cycle costs based on Deming's quality vs. cost model

In principle, the same argument applies to reliability: all efforts to improve reliability by identifying and removing potential causes of failures in service should result in cost savings later in the product life cycle, giving a net benefit in the longer term. In other words, an effective reliability programme represents an investment, usually with a large payback over a period of time. Unfortunately it is not easy to quantify the effects of given reliability programme activities, such as additional design analysis or testing, on achieved reliability. The costs (including those related to the effects on project schedules) of the activities are known, and they arise in the short term, but the benefits arise later and are often much less certain. However, achieving levels of reliability close to 100% is often not realistic for complex products. Recent research on reliability cost modelling showed that in practical applications the total cost curve is highly skewed to the right due to the increasing cost and diminishing return on further reliability improvements, as shown in Figure 1.8. The tight time scales and budgets of modern product development can also impact the amount of effort that can be applied. On the other hand there is often strong market pressure to achieve near perfect reliability.

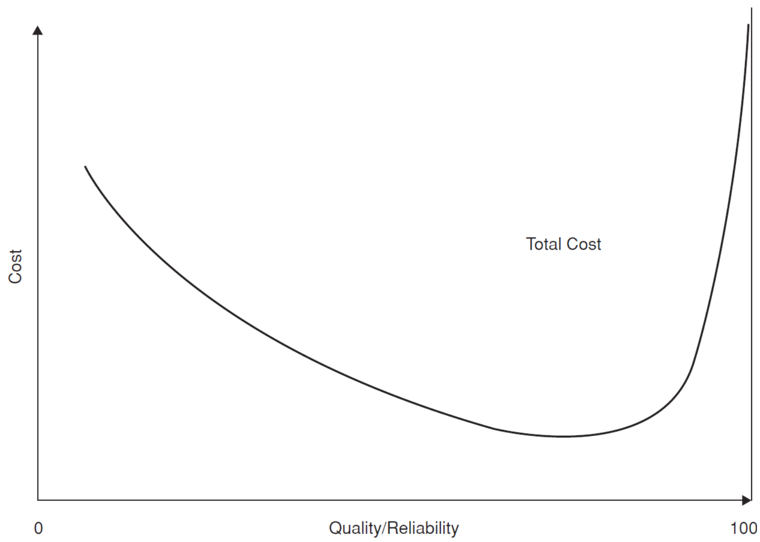


Figure 1.8: Reliability and life cycle costs (practical applications)

It is important to remember though that while achieving 100% quality in manufacturing operations, or 100% reliability in service, is extremely rare in real life applications, especially in high volume production, it should nevertheless be considered as an ultimate goal for any product development and production programme. Achieving reliable designs and products requires a totally integrated approach, including design, test, production, as well as the reliability programme activities. The integrated engineering approach places high requirements for judgment and engineering knowledge on project managers and team members. Reliability specialists must play their parts as members of the team. There are three kinds of engineering products, from the perspective of failure prevention:

1. **Intrinsically reliable components**, which are those that have high margins between their strength and the stresses that could cause failure, and which do not wear out within their practicable life times. Such items include nearly all electronic components (if properly applied), nearly all mechanical non-moving components, and all correct software.
2. **Intrinsically unreliable components**, which are those with low design margins or which wear out, such as badly applied components, light bulbs, turbine blades, parts that move in contact with others, like gears, bearings and power drive belts, and so on.
3. **Systems** which include many components and interfaces, like cars, dishwashers, aircraft, and so on, so that there are many possibilities for failures to occur, particularly across interfaces (e.g. inadequate electrical overstress protection, vibration nodes at weak points, electromagnetic interference, software that contains errors, and so on).

It is the task of design engineers to ensure that all components are correctly applied, that margins are adequate (particularly in relation to the possible extreme values of strength and stress, which are often variable), that wear-out failure modes are prevented during the expected life (by safe life design, maintenance, etc.), and that system interfaces cannot lead to failure (due to interactions, tolerance mismatches, etc.). Because achieving all this on any modern engineering product is a task that challenges the capabilities of the very best engineering teams, it is almost certain that aspects of the initial design will fall short of the "intrinsically reliable" criterion. Therefore we must submit the design to analyses and tests in order to show not only that it works, but also to show up the features that might lead to failures. When we find out what these are we must redesign and re-test, until the final design is considered to meet the criterion. Then the product has to be manufactured. In principle, every one should be identical and correctly made. Of course this is not achievable, because of the inherent variability of all manufacturing processes, whether performed by humans or by machines. It is the task of the manufacturing people to understand and control variation, and to implement inspections and tests that will identify non-conforming product. For many engineering products the quality of operation and maintenance also influence reliability. The essential points that arise from this brief and obvious discussion of failures are that:

1. Failures are caused primarily by people (designers, suppliers, assemblers, users, maintainers). Therefore the achievement of reliability is essentially a management task, to ensure that the right people, skills, teams and other resources are applied to prevent the creation of failures.
2. Reliability (and quality) specialists cannot by themselves effectively ensure the prevention of failures. High reliability and quality can be achieved only by effective team working by all involved.
3. There is no fundamental limit to the extent to which failures can be prevented. We can design and build for ever-increasing reliability.

Deming explained how, in the context of manufacturing quality, there is no point at which further improvement leads to higher costs. This is, of course, even more powerfully true when considered over the whole product life cycle, so that efforts to ensure that designs are intrinsically reliable, by good design, thorough analysis and effective development testing, can generate even higher pay-offs than improvements in production quality. The "kaizen" (continuous improvement) principle is even more effective when applied to up-front engineering. The creation of reliable products is, therefore, primarily a management task.

1.7 Chapter Questions

1. Define (a) failure rate, and (b) hazard rate. Explain their application to the reliability of components and repairable systems. Discuss the plausibility of the 'bathtub curve' in both contexts.
2. Explain the theory of component failures derived from the interaction of stress (or load) and strength distributions. Explain how this theory relates to the behaviour of the component hazard function.
3. Discuss the validity of the bathtub curve when used to describe the failure characteristics of non-repairable parts.
4. What are the main objectives of a reliability engineering team working on an engineering development project? Describe the important skills and experience that should be available within the team.
5. Briefly list the most common basic causes of failures of engineering products.
6. It is sometimes claimed that increasing quality and reliability beyond levels that have been achieved in the past is likely to be uneconomic, due to the costs of the actions that would be necessary. Present the argument against this belief. Illustrate it with an example from your own experience.
7. Describe the difference between repairable and non-repairable parts. What kind of effect might this difference have on reliability? List examples of repairable and non-repairable parts in your everyday life.
8. Explain the difference between reliability and durability and how they can be specified in a product development programme.
9. List the potential economic outcomes of poor reliability, and identify which costs are directly quantifiable and which are intangible. Explain how they can be minimised, and discuss the extent to which very high reliability (approaching zero failures) is achievable in practice.
10. What are the major factors that might limit the achievement of very high reliability?
11. After processing the existing programme cost data and running a regression model on the previous projects, the cost of product development and manufacturing (CDM) has been estimated to follow the equation: CDM $= 0.8$million rands $+ 3.83$million rands $\times R^2$ (R is the achieved product reliability at service life and is expected to be above 90%). The cost of failure (CF) has been estimated as the sum of fixed cost of 40 000 rands plus variable cost of 150 rands per failure. The total number of the expected failures is $n \times (1 - R)$, where n is the total number of produced units. Considering that the production volume is expected to be 50 000 units, estimate the optimal target reliability and the total cost of the programme. (Ans: R = 0.979, Total cost = R4 668 329.03)
12. Select an everyday item (coffee maker, lawnmower, bicycle, mobile phone, CD player, computer, refrigerator, microwave oven, cooking stove, etc.). (a) Discuss the ways this item can potentially fail. What can be done to prevent those failures? (b) Based on the Figures 1.2 and 1.3, what would be an example of the load and strength for a critical component within this item? Do you expect load and strength for this component to be time-dependent?

Chapter 2

Reliability Modelling

2.1 Probability Theory

It is required to review the basic concepts of probability theory to serve as basis for modelling the reliability of systems. It is important to note that only essential concepts to support reliability modelling are presented. For comprehensive literature and background about the concepts in this section refer to textbooks such as Walpole et al. (1998, 10-48), O'Connor and Kleyner (2012, 21-28) and McCool (2012, 1-11).

2.1.1 Probability Definitions

The following definitions applies to probability theory (Walpole et al., 1998, 11, 14, 15):

- The set of all possible outcomes of a statistical experiment is called the **sample space** and represented by S.
- For any experiment, there is interest in the occurrence of certain events. An **event** is a subset op a sample space typically denoted as A, B, C, ...
- The **complement** of an event A with respect to S is the subset of all elements of S that are not in A and is denoted by $\overline{A}$.
- The **intersection** of two events A and B, denoted by the symbol $A \cap B$, is the set of outcomes that belong to both A and B.
- Two events A and B are **mutually exclusive** if $A \cap B = \emptyset$, in other words if A and B have no elements in common.
- The **union** of the two events A and B, denoted by the symbol $A \cup B$, is the set of outcomes that belong to A or B or both.

2.1.2 Probability of an Event

The likelihood of the occurrence of an event resulting from a statistical experiment is evaluated by means of a set of real numbers called **probabilities** ranging from 0 to 1 (Walpole et al., 1998, 27).

To find the probability of an event A, we sum all the probabilities assigned to the sample points in A. This sum is call the probability of A and is denoted by $P(A)$, with the following attributes (Walpole et al., 1998, 28):

$$\begin{aligned} 0 \leq P(A) \leq 1 \\ P(A) = 1 - P(\overline{A}) \\ P(\emptyset) = 0 \\ P(S) = 1 \end{aligned} \tag{2.1}$$

It follows that event A will definitely occur, if it has a probability of 1, while it is impossible for event A to occur if its probability is equal to zero.

If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is (Walpole et al., 1998, 29):

$$P(A) = \frac{n}{N} \tag{2.2}$$

2.1.3 Additive Rules

There are several laws which simplifies the computation of probabilities. The first – the additive rule – applies to the union of events. The union of two events, A and B (Figure 2.1) is defined as (Walpole et al., 1998, 30):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2.3}$$

If A and B have no elements in common, or are mutually exclusive 2.3 reduces to (Walpole et al., 1998, 31):

$$P(A \cup B) = P(A) + P(B) \tag{2.4}$$

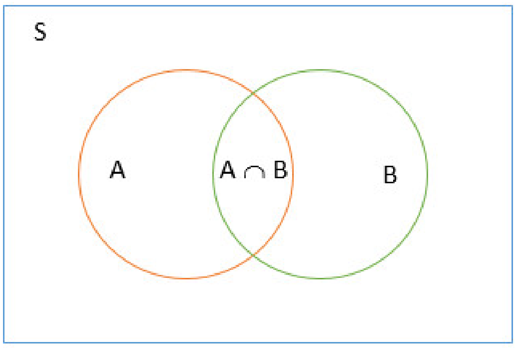


Figure 2.1: Venn diagram for the union of A and B

2.1.4 Conditional Probability

The probability of an event A occurring when it is known that some event B has occurred is called **conditional probability** and is defined as (Walpole et al., 1998, 35):

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad \text{if } P(A) > 0 \tag{2.5}$$

By multiplying 2.5 with $P(A)$ the multiplicative rule is obtained for calculating the intersection between A and B (Walpole et al., 1998, 38):

$$P(A \cap B) = P(B|A)P(A) \tag{2.6}$$

In the case where events A and B are *independent* of each other. In other words the occurrence of B has no impact on the odds of A occurring (Walpole et al., 1998, 37):

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B) \tag{2.7}$$

Then 2.6 becomes:

$$P(A \cap B) = P(B)P(A) \tag{2.8}$$

It follows that if, in an experiment, the independent events A_1, A_2,A_3, $\ldots$, A_k can occur their intersection is given by (Walpole et al., 1998, 40):

$$\begin{gathered} P(A_1 \cap A_2 \cap A_3 \cap \ldots A_k) = P(A_1)P(A_2)P(A_3)\ldots P(A_k) \\ P\left[\bigcap_{i=1}^{k} A_i\right] = \prod_{i=1}^{k} P(A_i) \end{gathered} \tag{2.9}$$

***Example* 2.1.** *Consider an experiment of tossing a die (Walpole et al., 1998, 14, 15). If we are interested in the number that shows on the top face, the sample space would be $S_1 = 1, 2, 3, 4, 5, 6$. Let A be the event of throwing an even number and B the event of throwing a number greater than 3. Then, $A = 2, 4, 6$ and $B = 4, 5, 6$. It follows that:*

$$\begin{aligned} A \cup B &= \{2, 4, 5, 6\} \\ A \cap B &= \{4, 6\} \\ \overline{A} &= \{1, 3, 5\} \end{aligned}$$

Suppose we are interested in event C, the probability of throwing a 2 on a single die roll. Event C is therefore represented as, $C = \{2\}$. The probability of event C occurring, according to 2.2 is then:

$$\begin{aligned} P(C) &= \frac{n}{N} \\ &= \frac{1}{6} \\ &= 0, 1\dot{6} \end{aligned}$$

For a second independent roll of the die the probability of rolling a 2 again is $P(C) = 0, 1\dot{6}$. However if we are interested in the probably of C occurring on two consecutive rolls 2.8 applies:

$$\begin{aligned} P(C \cap C) &= P(C)P(C) \\ &= \frac{1}{6} \times \frac{1}{6} \\ &= 0, 02\dot{7} \end{aligned}$$

2.2 Reliability of Systems

RBD and FTA are two commonly used symbolic, analytical techniques used for the analysis of system reliability and characteristics. Each has its own symbolic structure. FTA analyse failure combinations and is traditionally used to analyse systems with fixed probabilities. FTA can also easily be converted to RBD. RBD modelling analyse success combinations to describe the reliability of a systems. This technique is more flexible and allows for time-varying distrubtions for success and repair distributions. RBD are generally more difficult to convert to FTA. RBD shows the logical connections of components needed to fulfil a specified system function. If the system has more than one function, each function is considered individually, and a separate reliability block diagram has to be established for each system function. (Rausand and Hoyland, 2004, 118). The remainder of this chapter will cover the RBD technique.

2.2.1 Basic Series Systems

A system that is functioning if and only if all of its n components are functioning is called a *series* system. The corresponding reliability block diagram is shown in figure 2.2 (Rausand and Hoyland, 2004, 119).

Figure 2.2: Series system

Let A_i and $P(A_i)$ be the event and probability that component i is functioning. Since all n items need to function for the system to function successfully the reliability of the system is determined by the intersection of $P(A_i)$ according to 2.6 (Hines et al., 2003, 24):

$$\begin{aligned} R &= P(A_1 \cap A_2 \cap \cdots \cap A_n) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\dots P(A_n|A_1A_2\dots A_{n-1}) \end{aligned} \tag{2.10}$$

For the case where the events, A_i, are independent 2.10 is reduced according to 2.8 and 2.9 to:

$$\begin{aligned} R &= P(A_1)P(A_2)\dots P(A_n) \\ &= \prod_{i=1}^{n} P(A_i) \\ &= \prod_{i=1}^{n} R_i \end{aligned} \tag{2.11}$$

The reliability of a series system is thus equal to the product of the reliabilities of the individual items.

***Example* 2.2.** *Three components are reliability-wise in series and forms a system. Component 1 has a reliability of 99.5%, component 2 of 98.7% and component 3 of 97.3% for a mission of 100 hours. What is the overall reliability of the system for a 100-hour mission?*

Figure 2.3: A 3-component series system

$$\begin{aligned} R_S &= \prod_{i=1}^{n} R_i \\ &= R_1 \cdot R_2 \cdot R_3 \\ &= 0.995 \cdot 0.987 \cdot 0.973 \\ &= 0.9555 \end{aligned}$$

***Example* 2.3.** *Given a series system with five components, and a system mission reliability target of at least 0.99, determine the target component reliability levels, such that all components' targets are equal.*

$$\begin{aligned} R_S &= \prod_{i=1}^{n} R_i \\ &= R_i^5 = 0.99 \\ \Rightarrow R_i &= \sqrt[5]{0.99} = 0.998 \end{aligned}$$

The implication of 2.11 is that the least reliable component has the biggest effect on the system. The reliability of a series system is always less than the reliability of the least reliable component. A further consideration to keep in mind is that a series system's reliability decreases as the number of components in the system increases as seen in figure 2.4.

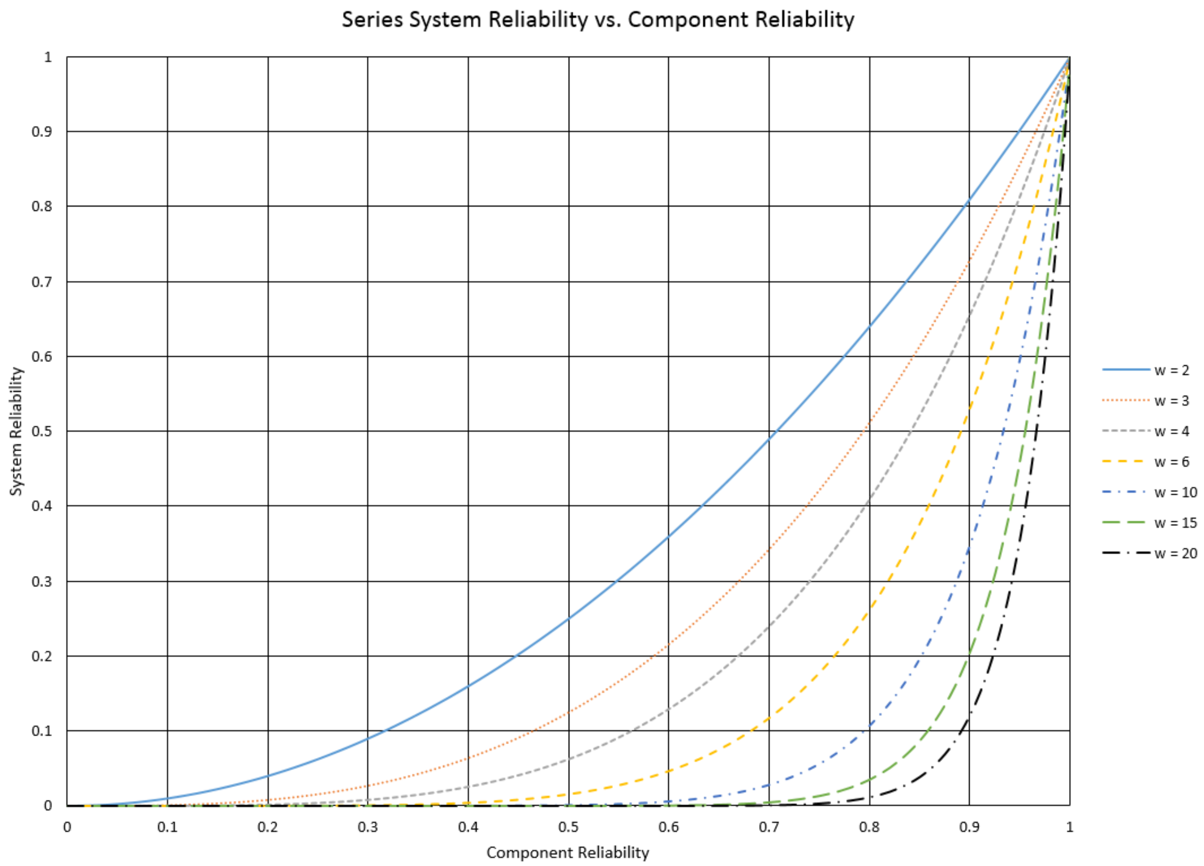


Figure 2.4: Series system reliability as a function of the individual component reliability

2.2.2 Redundancy Configurations

A system's reliability can be improved by introducing redundancy in the form of reserve components. When more than one component is placed to operate in parallel, it is referred to as *active redundancy*. However, reserves components can also be kept on standby and only activated when the primary component fails. If the reserve components carry no load before activation (and therefore cannot fail during this time) it is referred to as *passive redundancy* (Rausand and Hoyland, 2004, 173).

Active (Parallel) Redundancy

A system that is functioning if <u>at least one</u> of its n components is functioning is called an *active parallel* system. From this it follows that a parallel system therefore fails <u>if all</u> of its n components have failed. The corresponding reliability block diagram is shown in figure 2.5 (Rausand and Hoyland 2004, 119-120; O'Connor and Kleyner 2012, 21-28).

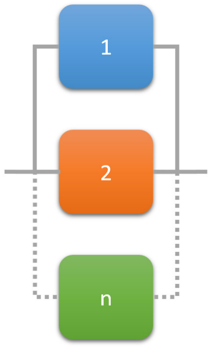


Figure 2.5: Parallel system

Let A_i and $P(A_i)$ be the event and probability that component i has failed. Since all n items need to fail for the system to fail the probability of system failure, F (also referred to as unreliability), is determined by the intersection of $P(A_i)$ according to 2.6 (Hines et al., 2003, 24):

$$\begin{aligned} F &= P(A_1 \cap A_2 \cap \cdots \cap A_n) \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1}) \end{aligned} \tag{2.12}$$

For the case where the events, A_i, are independent 2.12 is reduced according to 2.8 and 2.9 to:

$$\begin{aligned} F &= P(A_1)P(A_2)\ldots P(A_n) \\ &= \prod_{i=1}^{n} P(A_i) \\ &= \prod_{i=1}^{n} F_i \end{aligned} \tag{2.13}$$

From 2.1 it follows that $F = 1 - R$. Therefore the reliability of a parallel system follows from 2.13:

$$\begin{aligned} F &= \prod_{i=1}^{n} F_i \\ 1 - R &= \prod_{i=1}^{n} (1 - R_i) \\ R &= 1 - \prod_{i=1}^{n} (1 - R_i) \end{aligned} \tag{2.14}$$

***Example* 2.4.** *Three components are reliability-wise in parallel and forms a system. Component 1 has a reliability of 99.5%, component 2 of 98.7% and component 3 of 97.3% for a mission of 100 hours. What is the overall reliability of the system for a 100-hour mission?*

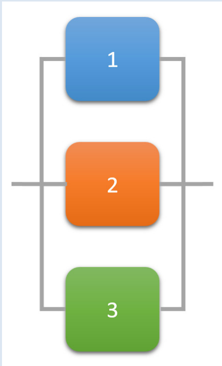


Figure 2.6: A 3-component series system

$$
\begin{aligned}
R_S &= 1 - \prod_{i=1}^{n}(1 - R_i) \\
&= 1 - (1 - R_1)(1 - R_2)(1 - R_3) \\
&= 1 - (1 - 0.995)(1 - 0.987)(1 - 973) \\
&= 1 - 0.00000175 \\
&= 0.999998245
\end{aligned}
$$

The implication of 2.14 is that the component with the highest reliability has the biggest effect on the system, since it is likely to fail last. A further consideration to keep in mind is that a parallel system's reliability increases as the number of components in the system increases as seen in figure 2.7.

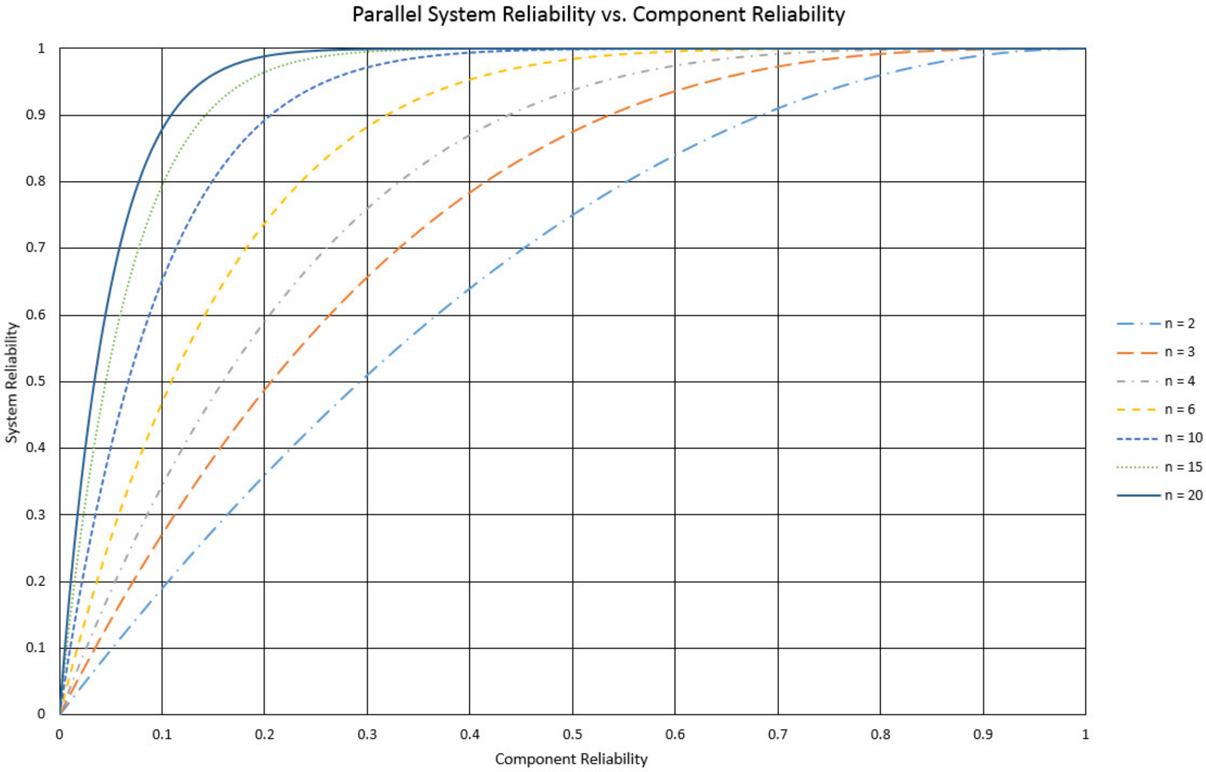


Figure 2.7: Parallel system reliability as a function of the individual component reliability

k-out-of-n Redundancy

In some cases systems are configured that survival of k items out of a total of n items are sufficient to ensure system survival (figure 2.8). This is a special case of parallel redundancy.

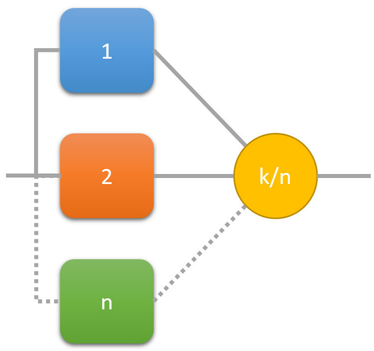


Figure 2.8: Representation of a 2-out-of-3 sytem

First consider a system consisting of n identical items, where R_i is constant for $i = 1, 2, \ldots, n$, and only k of these items are required for system survival. Using the binomial function, the system's reliability can be calculated as (Rausand and Hoyland, 2004, 151),

$$R = \sum_{j=k}^{n} \binom{n}{k} R^j (1-R)^{n-j}$$
$$\text{where} \binom{n}{k} = \frac{n!}{j!(n-j)!} \tag{2.15}$$

If reliabilities are difference for each component, it is not possible to make use of the binomial function and each combination of the k items that will ensure system survaval needs to be identified. The general for of the system reliability is given by:

$$R = \left[\sum_{l} \left(\prod_{i} R_i \prod_{j} F_j \right) \right] + \prod_{y=1}^{n} R_y \tag{2.16}$$

where l counts to the total number of possible combinations, i indicates the items required to survive in any combination and j indicates the items allowed to fail in any combination.

***Example* 2.5.** *Consider a system of 6 pumps of which at least 4 must function for the system success. Each pump has an 85% reliability for the operational duration. What is the probability of system success for the operational duration?*

$$\begin{aligned} R_s &= \sum_{j=4}^{6} \binom{6}{j} 0.85^j (1-0.85)^{6-j} \\ &= \binom{6}{4} 0.85^4 (1-0.85)^2 + \binom{6}{5} 0.85^5 (1-0.85)^5 + \binom{6}{6} 0.85^6 (1-0.85)^0 \\ &= 0.1762 + 0.3993 + 0.3771 \\ &= 95.26\% \end{aligned}$$

***Example* 2.6.** *Three hard drive in a computer system are configure in parallel. At least two of the hard drives must function in order fo the computer to work properly. Each hard drive is the same size and speed, but they are made by different manufacturers and have different reliabilities. The reliability of hard drive #1 is $R_1 = 0.9$, of hard drive #2 is $R_2 = 0.88$ and of hard drive #3 is $R_3 = 0.85$, all at the same mission time. The calculate the system reliability the following operational combinations are possible:*

	HD 1	*HD 2*	*HD 3*
1	*Failed*	*Operational*	*Operational*
2	*Operational*	*Failed*	*Operational*
3	*Operational*	*Operational*	*Failed*
4	*Operational*	*Operational*	*Operational*

The system reliability according to equation 2.16 is given by:

$$\begin{aligned} R_s &= R_2 R_3 F_1 + R_1 R_3 F_2 + R_1 R_2 F_3 + R_1 R_2 R_3 \\ &= 0.88 \cdot 0.85 \cdot 0.1 + 0.9 \cdot 0.12 \cdot 0.85 + 0.9 \cdot 0.88 \cdot 0.15 + 0.9 \cdot 0.88 \cdot 0.85 \\ &= 0.9586 \\ &= 95.86\% \end{aligned}$$

Passive (Standby) Redundancy

Passive redundancy is often referred to as "cold standby" and describes the situation where one component does not operate continuously, but is only activated when the primary component fails (figure 2.9). Modelling passive redundancy systems is more complex because the time for which the standby component is required to operate is a variable.

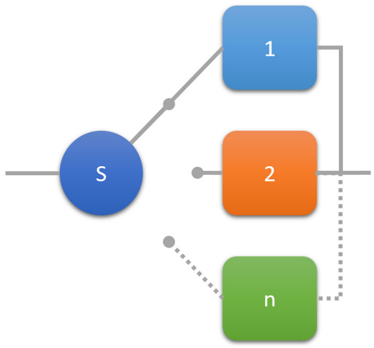


Figure 2.9: Representation of a passive redundancy configuration

For a two component ($n = 2$) passive redundancy configuration the system reliability is determined by (Sherwin and Bossche, 1993, 51):

$$\begin{aligned} R &= R_1(t) + \int_0^t \Big[f_1(x) \int_{t-x}^{\infty} f_2(u)du \Big] dx \\ &= R_1(t) + \int_0^t [f_1(x) R_2(t-x)] dx \end{aligned} \tag{2.17}$$

However, standby components are usually associated with complex system, rather than single components. Such systems tend to have constant ROCOF, $\lambda = \frac{1}{\text{MTBF}}$ under an unchanging maintenance policy during the majority of its live. **In the case of a constant failure rate (CFR) and identical components** it is legitimate to assume a Poisson process, with parameter $m = \lambda t$ (Sherwin and Bossche, 1993, 51). Therefore, for a $n = 2$ system, 2.17 becomes:

$$R = (1 + \lambda t)e^{-\lambda t} \tag{2.18}$$

The general function for n identical components in a standby redundant configuration, with perfect switching is:

$$\begin{aligned} R &= e^{-\lambda t} \sum_{i=0}^{n-1} \frac{(\lambda t)^i}{i!} \\ &= 1 - \frac{(-\lambda t)^n e^{-\lambda t}}{n!} \end{aligned} \tag{2.19}$$

For a k-out-of-n system the total failure rate is $k\lambda$ throughout and the system is stopped after $n - k + 1$ failures, therefore 2.19 becomes:

$$R = e^{-k\lambda t} \sum_{i=0}^{n-k} \frac{(k\lambda t)^i}{i!} \tag{2.20}$$

Refer to Rausand and Hoyland (2004, 175-177) for the modelling of imperfect switching and partly loaded redundancy scenarios.

***Example* 2.7.** *A shipboard missile system consists of two warning radars, a launch guidance system and the missiles. The radars are arranged so that either can give warning if the other fails, in a standby redundant configuration with perfect switching. Four missiles are available for firing and the system is considered to be reliable if three out of the four missiles can be fired and guided. Figure 2.10 shows the RBD of the system. Over a 24hr period the reliability of each missile is 0.9 and for the launch guidance system it is 0.9685. The MTBF of both the primary and secondary radar systems is 1000 hours and 750 hours for the launch guidance system. Determine the reliability of the system over 24hr (O'Connor and Kleyner, 2012, 146).*

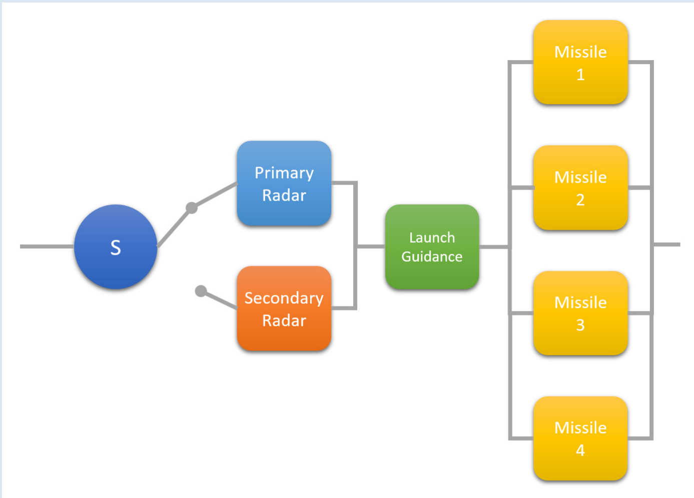


Figure 2.10: Illustration of the missile system

$$\begin{aligned}
R_R &= (1+\lambda t)e^{-\lambda t} \\
&= (1+\frac{1}{1000}\cdot 24)e^{-\frac{1}{1000}\cdot 24} \\
&= 0.9997 \\
R_M &= \sum_{j=k}^{n}\binom{n}{k}R_M^j(1-R_M)^{n-j} \\
&= \binom{4}{3}R_M^3(1-R_M)^1+\binom{4}{4}R_M^4(1-R_M)^0 \\
&= \binom{4}{3}0.9^3(1-0.9)^1+\binom{4}{4}0.9^4(1-0.9)^0 \\
&= 0.9477 \\
R_S &= R_R R_{LG} R_M \\
&= 0.9997\cdot 0.9685\cdot 0.9477 \\
&= 0.9176
\end{aligned}$$

2

2.2.3 Complex Configurations and Analysis

4 System reliability configuration rarely consist of only series or redundancy arrangements. Normally RBD are more complex consisting of combinations of these arrangements.

Modular Decomposition

In modular decomposition the block diagram analysis consist of reducing the overall RBD to a simple system which can then be analysed using the formulae for series and redundancy arrangements (Rausand and Hoyland 2004, 138; O'Connor and Kleyner 2012, 152). Consider the system in figure 2.11. Component 1 has a reliability of 99.5%, component 2 of 98.7% and component 3 of 97.3% for a mission of 100 hours. What is the overall reliability of the system for a 100-hour mission?

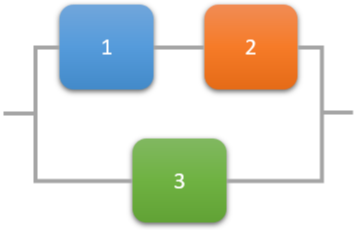


Figure 2.11: A RBD consisting of a series and redundancy arrangement

$$
\begin{aligned}
R_{12} &= R_1 R_2 \\
&= 0.995 \cdot 0.987 \\
&= 0.982 \\
R_S &= 1-(1-R_{12})(1-R_3) \\
&= 1-(1-0.982)(1-0.973) \\
&= 0.9995
\end{aligned}
$$

***Example* 2.8.** *Consider the complex system with series and redundancy arrangements in figure 2.12. The component reliabilities are as follows:* $R_1 = 0.983$, $R_2 = 0.995$, $R_3 = 0.963$, $R_4 = 0.995$, $R_5 = 0.955$, $R_6 = 0.912$, $R_7 = 0.924$, $R_8 = 0.999$, $R_9 = 0.991$, $R_{10} = 0.998$ *and* $R_{11} = 0.938$. *Determine the systems reliability.*

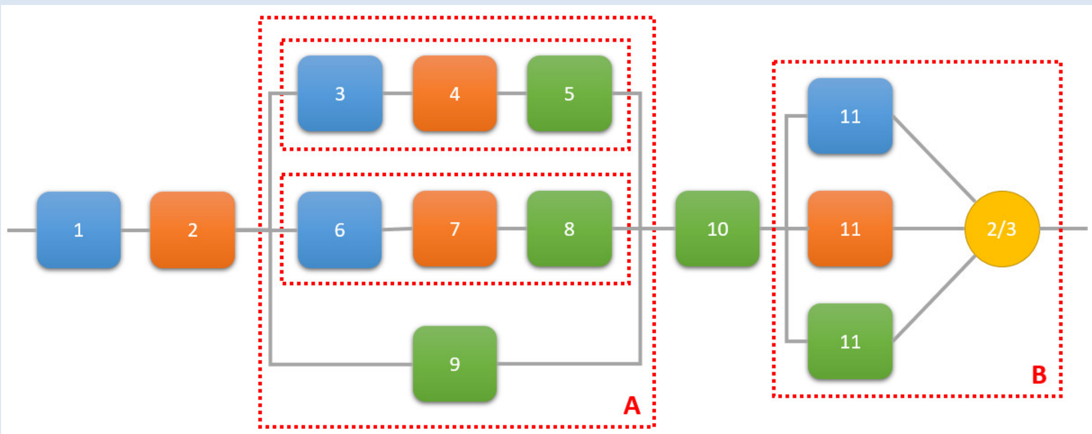


Figure 2.12: Complex system with multiple series and redundancy arrangements

This system can be reduced as follows:

$$\begin{aligned}
R_A &= 1 - (1 - R_3R_4R_5)(1 - R_6R_7R_8)(1 - R_9) \\
&= 1 - (1 - 0.963 \cdot 0.995 \cdot 0.955)(1 - 0.912 \cdot 0.924 \cdot 0.999)(1 - 0.991) \\
&= 0.9999 \\
R_B &= \sum_{j=k}^{n} \binom{n}{k} R_{11}^{j}(1 - R_{11})^{n-j} \\
&= \binom{3}{2} 0.938^2(1 - 0.938)^1 + \binom{3}{3} 0.938^3(1 - 0.938)^0 \\
&= 0.9889 \\
R_S &= R_1R_2R_AR_{10}R_B \\
&= 0.983 \cdot 0.995 \cdot 0.9999 \cdot 0.998 \cdot 0.9889 \\
&= 0.9652
\end{aligned}$$

2.3 Component Importance

Once the reliability of system has been modelled, it is often required to identify the least reliable components of the system to improve the design. A component in series is generally more important than a component in a redundancy arrangement. In reliability engineering there are various importance measures available for sensitivity analysis. These measures can be used to rank or classify components according to their importance (Rausand and Hoyland, 2004, 183-9).

The six most used importance measures are: Birnbaum's measure; improvement potential measure; risk improvement worth; risk reduction worth; criticality importance measure; and Fussel-Vesely's measure. Each of these measures are based on slightly different interpretations of component importance. All however take the component location and its reliability into account.

The objective of component importance measurement is therefore to help the designer to identify components that should be improved, while during the operational phase of the systems life it is useful for allocating inspection and maintenance resources to the most important components. Birnbaum's measure is the most widely used component importance measure and will be discussed in further detail. Refer to Rausand and Hoyland (2004, 189-97) for details about other component importance measures.

Birnbaum (1968) proposed the following measure of importance of component i, at time, t:

$$I_i = \frac{\partial R_S}{\partial R_i} \tag{2.21}$$

Birnbaum's measure is obtained by partial differentiation of the system reliability, R_S, with respect to the component's reliablity, R_i. This approach is well known in sensitivity analysis, where it follows that if I_i is large, a small change in the component's reliability will result in an equally large change in the system's reliability.

Before applying Birnbaum's measure, it could be appropriate to recap on some differential calculus. The most used differentiation rules relevant for determining component importance are the (Finney and Thomas, 1993, 126):

26 **Rule 1.** *If c is a constant, then*

$$\frac{\partial}{\partial x}(c) = 0 \tag{2.22}$$

Rule 2. *If n is a positive or negative integer, then*

$$\frac{\partial}{\partial x}(x^n) = nx^{n-1} \tag{2.23}$$

Rule 3. *If u and v are differentiable functions of x, then*

$$\frac{\partial}{\partial x}(u+v) = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x} \tag{2.24}$$

2

$$\frac{\partial}{\partial x}(u-v) = \frac{\partial u}{\partial x} - \frac{\partial v}{\partial x} \tag{2.25}$$

***Example* 2.9.** *Consider a simple series system with $n = 2$, where $R_1 = 0.98$ and $R_2 = 0.96$. The system's reliability at time, t is:*

$$R_S = R_1 R_2$$

By applying 2.23 the Birnbaum's measure for the components can be obtained:

$$\begin{aligned} I_1 &= \frac{\partial R_S}{\partial R_1} \\ &= 1 \cdot R_1^{1-1} \cdot R_2 \\ &= R_2 \\ &= 0.96 \\ I_2 &= \frac{\partial R_S}{\partial R_2} \\ &= R_1 \cdot 11 \cdot R_2^{1-1} \\ &= R_1 \\ &= 0.98 \end{aligned}$$

It therefore follows that the component with the lowest reliability in a series system is the most important.

***Example* 2.10.** *Now consider a simple parallel system with $n = 2$, where $R_1 = 0.98$ and $R_2 = 0.96$. The system's reliability at time, t is:*

$$\begin{aligned} R_S &= 1 - (1 - R_1)(1 - R_2) \\ &= 1 - (1 - R_2 - R_1 + R_1 R_2) \\ &= R_2 + R_1 - R_1 R_2 \end{aligned}$$

By applying 2.22 and 2.23 the Birnbaum's measure for the components can be obtained:

$$\begin{aligned}
I_1 &= \frac{\partial R_S}{\partial R_1} \\
&= 0 + 1 \cdot R_1^{1-1} - 1 \cdot R_1^{1-1} \cdot R_2 \\
&= 1 - R_2 \\
&= 0.04 \\
I_2 &= \frac{\partial R_S}{\partial R_2} \\
&= 1 \cdot R_2^{1-1} + 0 - 1 \cdot R_2^{1-1} \cdot R_1 \\
&= 1 - R_1 \\
&= 0.02
\end{aligned}$$

It therefore follows that the component with the highest reliability in a parallel system is the most important.

***Example* 2.11.** *Consider a 2-out-of-3 system, where $R_1 = 0.98$, $R_2 = 0.96$ and $R_3 = 0.94$. The system's reliability at time, t is:*

$$\begin{aligned}
R_S &= R_1R_2(1-R_3) + R_1R_3(1-R_2) + R_2R_3(1-R_1) + R_1R_2R_3 \\
&= R_1R_2 - R_1R_2R_3 + R_1R_3 - R_1R_2R_3 + R_2R_3 + R_1R_2R_3 + R_1R_2R_3 \\
&= R_1R_2 + R_1R_3 + R_2R_3 - 2R_1R_2R_3
\end{aligned}$$

Birnbaum's measure for the components can now be obtained:

$$\begin{aligned}
I_1 &= \frac{\partial R_S}{\partial R_1} = R_2 + R_3 - 2R_2R_3 = 0.0952 \\
I_2 &= \frac{\partial R_S}{\partial R_2} = R_1 + R_3 - 2R_1R_3 = 0.0776 \\
I_3 &= \frac{\partial R_S}{\partial R_3} = R_1 + R_2 - 2R_1R_2 = 0.0584
\end{aligned}$$

2

***Example* 2.12.** *Consider the non-repairable system in figure 2.13. The component reliabilities are as follows: $R_1 = 0.97$, $R_2 = R_3 = 0.96$, $R_4 = R_{11} = 0.94$, $R_5 = 0.92$, $R_6 = 0.95$, $R_7 = 0.959$, $R_8 = 0.9$, $R_9 = 0.91$, $R_{10} = 0.93$ and $R_{12} = 0.99$. The system's reliability is provided. Proof it is correct and then determine the reliability importance of component 8.*

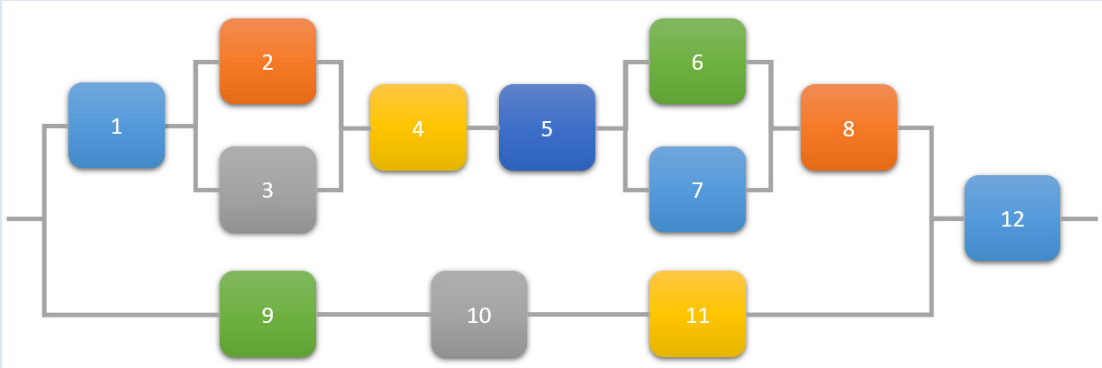


Figure 2.13: Complex system with multiple series and redundancy arrangements

This system's reliability is:

$$
\begin{aligned}
R_S = [&R_1(R_2 + R_3 + R_2R_3)R_4R_5(R_6 + R_7 - R_6R_7)R_8 + R_9R_{10}R_{11} \\
&-R_1(R_2 + R_3 - R_2R_3)R_4R_5(R_6 + R_7 - R_6R_7)R_8R_9R_{10}R_{11}]R_{12}
\end{aligned}
$$

Birnbaum's measure for component 8 is obtained:

$$
\begin{aligned}
I_8 &= \frac{\partial R_S}{\partial R_8} \\
&= R_1(R_2 + R_3 + R_2R_3)R_4R_5(R_6 + R_7 - R_6R_7)R_{12} \\
&= R_1(R_2 + R_3 - R_2R_3)R_4R_5(R_6 + R_7 - R_6R_7)R_9R_{10}R_{11}R_{12} \\
&= 0.97(0.96 + 0.96 - 0.96 \cdot 0.96) \cdot 0.94 \cdot 0.92(0.95 + 0.959 - 0.95 \cdot 0.959) \cdot 0.99 \\
&= -0.97(0.96 + 0.96 - 0.96 \cdot 0.96) \cdot 0.94 \cdot 0.92(0.95 + 0.959 - 0.95 \cdot 0.959) \cdot 0.91 \cdot 0.93 \cdot 0.94 \cdot 0.99 \\
&= 0.169
\end{aligned}
$$

4

Chapter 3

Reliability Data Analysis

Survival analysis is a branch of statistics for analysing the expected duration of time until one or more events happen. These events, for example can be the development of a disease, death or failure of component or system. This topic is known as **reliability theory** or **reliability analysis** in engineering, duration analysis in economics and event history analysis in sociology (Lee and Wang, 2013, xi).

Reliability analysis are characterised by uncertainty. For example, data may show that a certain pump fails at a constant failure rate (CFR) of 1000 hours. If a hundred of these pumps are installed and operated for 100 hours, it is uncertain whether any of the pumps will fail withing this time. A statement about the *probability* of failure, within specified statistical *confidence limits* is however possible (O'Connor and Kleyner, 2012, 19).

3.1 Functions and Terminology

In reliability analysis there are four functions which are of significant importance. The probability density function (PDF), denoted by $f(t)$, the cumulative probability density function or failure distribution function, denoted by $F(t)$, the reliability function, denoted by $R(t)$, and the hazard function, denoted by $h(t)$ (Coetzee, 2004, 401-9).

3.1.1 Probability Density Function

The PDF represents the probability of failure occurring at any specific time. At any given time the PDF gives the probability of a failure occuring during the following time unit. The PDF is given by:

$$f(t) = \frac{1}{N} \cdot \frac{\Delta n}{\Delta t}, \tag{3.1}$$

where Δn = number of failures in the time interval $[t, t + \Delta n]$, Δt = length of time interval, and N = original population.

If $\Delta t \to 0$ then,

$$f(t) = \frac{1}{N} \cdot \frac{dn}{dt} \tag{3.2}$$

Figures 3.1a and 3.2a show examples of a typical probability density function.

3.1.2 Failure Distribution Function

The failure distribution function or cumulative probability density (CPD) function gives the cumulative probability of failure or the probability that failure has occurred on or before a specific time. The failure distribution function further states that all components will ultimately fail, where $F(\infty) = 1)$. The function is given by:

$$F(t) = \frac{\sum n_i}{N}, \tag{3.3}$$

where $\sum n_i$ = number of failures up to time t, and N = original population.

If $\Delta t \rightarrow 0$ then,

$$F(t) = \int_0^t F(t)dt \tag{3.4}$$

Figures 3.1b and 3.2b show examples of a typical failure distribution function.

3.1.3 Reliability (Survival) Function

The reliability function is the complement of the failure distribution function, since it gives the probability of survival up to any specific time. The reliability function is given by:

$$R(t) = 1 - F(t) \tag{3.5}$$

Figures 3.1c and 3.2c show examples of a typical reliability function.

3.1.4 Hazard Function

The hazard rate function is important in reliability analysis, since the shape is indicative of the type of maintenance strategy which needs to be used to maintain the specific component. The hazard rate function is given by:

$$h(t) = \frac{1}{n(t)} \cdot \frac{\Delta n}{\Delta t}, \tag{3.6}$$

where Δn = number of failures in the time interval $[t, t + \Delta n]$, Δt = width of time interval, and $n(t)$ = population surviving at t.

Figures 3.1d and 3.2c show examples of a typical hazard rate function. The hazard rate is an indication of the risk of failure at a specific time. The hazard rate can take the form of three different shapes: a decreasing, constant or increasing hazard rate. The combined effect of the different hazard rates produces the so-called *bathtub curve*. An increasing hazard rate lends itself to a usage-based maintenance or replacement strategy, to reduce the hazard rate. In cases of a constant or decreasing hazard rate condition-based, run-to-failure or redesign are the only maintenance strategy options. Hazard rates are further discussed in §3.7.3. The hazard rate can also be translated into a relationship between the probability density and reliability functions, as shown in proof 3.7:

$$\begin{aligned}
h(t) &= \frac{1}{n(t)} \cdot \frac{\Delta n}{\Delta t} \\
&= \frac{Nf(t)}{n(t)} \\
&= \frac{Nf(t)}{N - \sum n_i} \\
&= \frac{f(t)}{\frac{N - \sum n_i}{N}} \\
&= \frac{f(t)}{1 - F(t)} \\
&= \frac{f(t)}{R(t)}
\end{aligned} \tag{3.7}$$

***Example* 3.1.** *Table 3.1 consists of failure observations, i, (in column 1) and operating times for 10 electronic components which failed in the same position (in column 2). The first component failed at 8 hours. It was replaced with a second new component, which then failed after another 12 hours, and so forth.*

Failure observation, i	***Operating time***
1	*8*
2	*20*
3	*34*
4	*46*
5	*63*
6	*86*
7	*111*
8	*141*
9	*186*
10	*266*

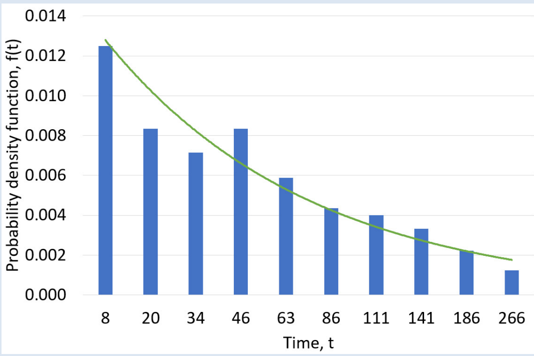
Table 3.1: *Typical failure data for an electronic component*

Table 3.2 shows the results of the calculations for the reliability functions 3.1, 3.3, 3.5 and 3.6, while figure 3.1 shows the respective plots. The following example shows the calculations for $i = 4$.

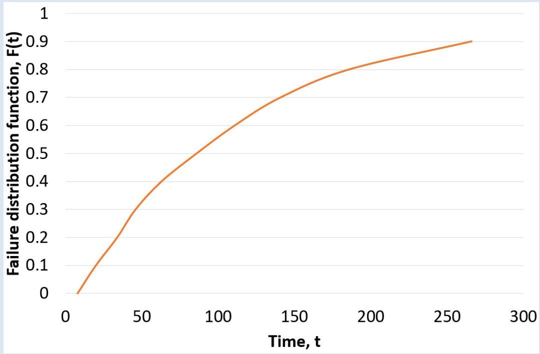
$$\begin{aligned}
f_4(t) &= \frac{1}{N} \cdot \frac{\Delta n}{\Delta t} = \frac{1}{10} \cdot \frac{1}{12} = 0.0083 \\
F_4(t) &= \frac{\sum n_i}{N} = \frac{3}{10} = 0.3 \\
R_4(t) &= 1 - F(t) = 1 - 0.3 = 0.7 \\
h_4(t) &= \frac{f(t)}{R(t)} = 0.0119
\end{aligned}$$

***Obs.*, i**	***Interval*, h**	***Operating time***	**Δt**	**f(t)**	**F(t)**	**R(t)**	**h(t)**
1	$0 \leq h < 8$	*8*	*8*	*0.0125*	*0*	*1*	*0.0125*
2	$8 \leq h < 20$	*20*	*12*	*0.0083*	*0.1*	*0.9*	*0.0093*
3	$20 \leq h < 34$	*34*	*14*	*0.0071*	*0.2*	*0.8*	*0.0089*
4	$34 \leq h < 46$	*46*	*12*	*0.0083*	*0.3*	*0.7*	*0.0119*
5	$46 \leq h < 63$	*63*	*17*	*0.0059*	*0.4*	*0.6*	*0.0098*
6	$63 \leq h < 86$	*86*	*23*	*0.0043*	*0.5*	*0.5*	*0.0087*
7	$86 \leq h < 111$	*111*	*25*	*0.0040*	*0.6*	*0.4*	*0.01*
8	$111 \leq h < 141$	*141*	*30*	*0.0033*	*0.7*	*0.3*	*0.0111*
9	$141 \leq h < 186$	*186*	*45*	*0.0022*	*0.8*	*0.2*	*0.0111*
10	$186 \leq h < 266$	*266*	*80*	*0.0013*	*0.9*	*0.1*	*0.0125*

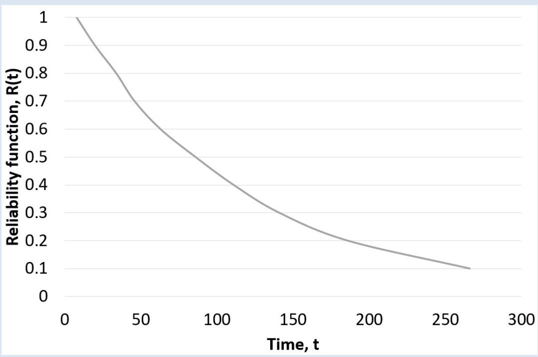
Table 3.2: *Results for the electronic component reliability functions*



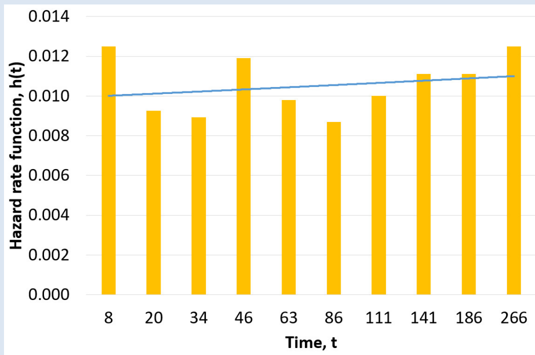
(a) Probability density function, $f(x)$



(b) Failure distribution function, $F(x)$



(c) Reliability function, $R(x)$



(d) Hazard rate function, $h(x)$

Figure 3.1: Results of reliability calculation for electronic component data

***Example* 3.2.** *Consider the case where 48 pumps are installed simultaneously and subject to the same operating conditions. Table 3.3 consists of the pump failure frequency (failure observations),* i*, (in column 1) and the respective time intervals over 12 months during which the pumps failed (in column 2). Two pumps failed during January, five during February, and so forth. By November all the pumps have failed.*

6

Failure Frequency, $\mathbf{i}$	***Time Interval***
2	*January*
5	*February*
7	*March*
8	*April*
7	*May*
6	*June*
5	*July*
4	*August*
3	*September*
1	*October*
0	*November*
0	*December*

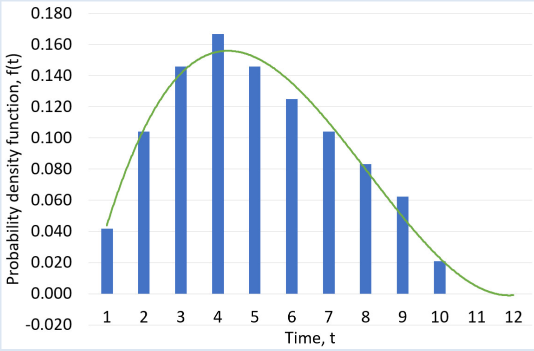
Table 3.3: *Typical failure data for pumps*

Table 3.4 shows the results of the calculations for the reliability functions 3.1, 3.3, 3.5 and 3.6, while figure 3.2 shows the respective plots. The following example shows the calculations for April (month 4).

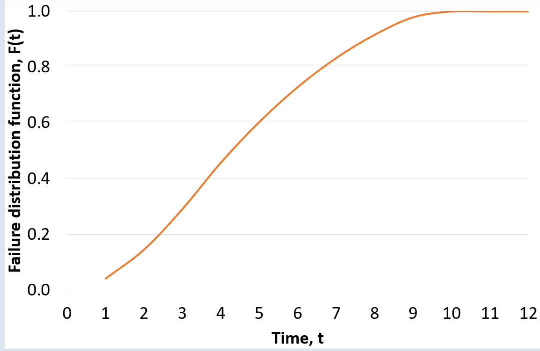
$$f_4(t) = \frac{1}{N} \cdot \frac{\Delta n}{\Delta t} = \frac{1}{48} \cdot \frac{8}{1} = 0.1667$$
$$F_4(t) = \frac{\sum n_i}{N} = \frac{22}{48} = 0.4583$$
$$R_4(t) = 1 - F(t) = 1 - 0.45833 = 0.5417$$
$$h_4(t) = \frac{f(t)}{R(t)} = 0.3077$$

Obs., $\mathbf{i}$	***Interval,*** $\mathbf{h}$	***Operating time***	$\mathbf{\Delta t}$	$\mathbf{f(t)}$	$\mathbf{F(t)}$	$\mathbf{R(t)}$	$\mathbf{h(t)}$
2	$0 \leq h < 1$	*1*	*1*	*0.0417*	*0.0417*	*0.9583*	*0.0435*
5	$1 \leq h < 2$	*2*	*1*	*0.1042*	*0.1458*	*0.8542*	*0.1220*
7	$2 \leq h < 3$	*3*	*1*	*0.1458*	*0.2917*	*0.7083*	*0.2059*
8	$3 \leq h < 4$	*4*	*1*	*0.1667*	*0.4583*	*0.5417*	*0.3077*
7	$4 \leq h < 5$	*5*	*1*	*0.1458*	*0.6042*	*0.3958*	*0.3684*
6	$5 \leq h < 6$	*6*	*1*	*0.1250*	*0.7292*	*0.2708*	*0.4615*
5	$6 \leq h < 7$	*7*	*1*	*0.1042*	*0.8333*	*0.1667*	*0.6250*
4	$7 \leq h < 8$	*8*	*1*	*0.0833*	*0.9167*	*0.0833*	*1*
3	$8 \leq h < 9$	*9*	*1*	*0.0625*	*0.9792*	*0.0208*	*3*
1	$9 \leq h < 10$	*10*	*1*	*0.0208*	*1*	*0*	*–*
0	$10 \leq h < 11$	*11*	*1*	*0*	*1*	*0*	*–*
0	$11 \leq h < 12$	*12*	*1*	*0*	*1*	*0*	*–*

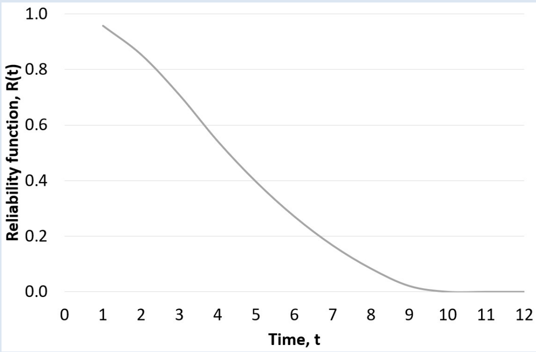
Table 3.4: *Results for the pump reliability functions*



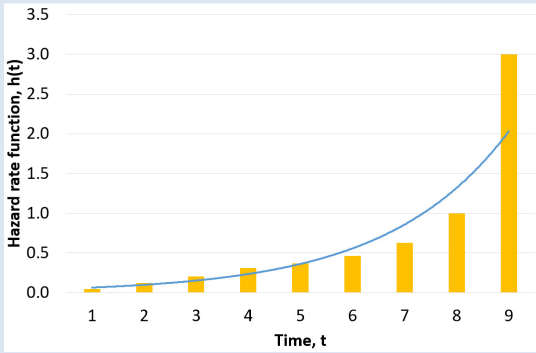
(a) Probability density function, $f(x)$



(b) Failure distribution function, $F(x)$



(c) Reliability function, $R(x)$



(d) Hazard rate function, $h(x)$

Figure 3.2: Results of reliability calculation for pump data

3.1.5 Terminology

The following terminology refers to different types of equipment and its components (Ascher and Feingold, 1984, 7):

Part An item which is not subject to disassembly and which is discarded after its first, and only failure.

Socket A space or equipment position which, at any given time, holds a part of a given type.

System A collection of two or more sockets and their associated parts, interconnected to perform one or more functions. All the parts of a system do not necessarily need to perform satisfactory in order for the system to perform satisfactory.

Nonrepairable System A system which is discarded the first time that it ceases to perform to specification.

Repairable System A system, which, after failing to perform at least one of its require functions, can be restored to performing all of its required functions by any method, other than replacement of the entire system.

It is further important to understand the mathematical foundation of reliability statistics. *Discrete* and *continuous probability distributions*, as well as *stochastic point processes*, are used in reliability statistics. Discrete distributions describe situations with discrete functions, such as a light bulb which either works or not, or a pressure vessel which passes or fails an inspection. The binomial and Poisson distributions are typically used for analysing these situations. Continuous distributions describe situations which are related to a continuous variable, such as time or distance travelled. The normal, lognormal, exponential, gamma and Weibull distributions are typically used for analysing these situations[1]. Stochastic point processes describe situations where more than one failure can occur in the time continuum. The ordinary renewal process, non-homogeneous Poisson process (NHPP) and generalised renewal process are typically used for analysing these processes. (O'Connor and Kleyner, 2012, 19)

3.2 Non-Repairable and Repairable Systems

It is important to distinguish between repairable and non-repairable items when predicting or measuring reliability. From this section and §3.1 the difference between non-repairable parts and repairable systems due to the fundamentally different statistical models which are used for analysing the life times of these systems should become clear.

3.2.1 Non-Repairable Systems

In reliability studies the domain of *life data analysis* refers to the study and modelling of **non-repairable parts**' lives. Visualise a socket into which a component is inserted at time zero. When the component fails, it is replaced immediately with a new one of the same kind. After each replacement , the socket is put back into an *as-good-as-new* condition. Each component has a time-to-failure that is determined by the underlying distribution. Discrete and continuous statistical functions (§3.1) form the mathematical basis of life data analysis, where the objective is to determine which distribution best fits time-to-failure data and for deriving estimates of the distribution parameters (O'Connor and Kleyner 2012, 70; Reliawiki.org 2016).

For a **non-repairable part** such as a light bulb, a transistor, a rocket motor or an unmanned spacecraft, **reliability is the survival probability over the item's expected life, or for a period during its life, when only one failure can occur**. It is characterised by the hazard rate (refer to §3.3.1) and life values such as the mean life or mean time to failure (MTTF), or the expected life by which a certain percentage might have failed (say 10%.) (percentile life), are other reliability characteristics that can be used. Note that non-repairable parts may be individual parts (light bulbs, transistors, fasteners) or systems comprised of many parts (spacecraft, microprocessors) (O'Connor and Kleyner, 2012, check).

Since a non-repairable part is discarded after failure, it means that one system's failure is independent from the next system's failure. Therefore, in life data analysis the events that are observed are assumed to be statistically independent and identically distributed (IID). A sequence or collection of random variables is said to be IID if:

- Each has the same probability distribution as any of the others.
- All are mutually independent, which implies that knowing whether or not one occurred makes it neither more nor less probable that the other occurred.

[1]It should be noted that the distinction between *discrete* and *continuous* functions depend on how the problem is treated and not about physics or mechanics. For example, the failure of the same vessel above could be a function of age and its reliability could be treated as a continuous function

Reliability literature also refers to non-repairable part modelling as *renewal theory*. Renewal theory is based on the principle that a part is replaced after failure to an as-good-as-new condition and that failures are IID. Ascher and Feingold (1984, 59) regard the renewal model as *poor*, since most repairs involve the replacement of only a proportion of the parts after a failure. Furthermore renewal theory is not limited to only non-repairable parts, because even if a system can physically be repaired it can still produce failure data that is IID and can therefore be classified as non-repairable.

3.2.2 Repairable Systems

In life data analysis, the part or component placed on test is assumed to be *as-good-as-new*. However, this is not the case when dealing with repairable systems that have more than one life.

Suppose that a system consists of many parts with each part in a socket. A failure in any socket constitutes a failure of the system. Each part in a socket is a non-repairable part (renewal process) governed by its respective distribution function. When the system fails due to a failure in a socket, the part is replaced and the socket is again *as-good-as-new*. The system has been repaired. Since there are many other parts still operating with various ages, the system is not typically *as-good-as-new*, but rather *same-as-old*. For example, a car is not *as-good-as-new* after the replacement of a failed water pump. For repairable systems analysis the events that are observed are part of a stochastic point process (§3.1). A stochastic process is defined as a sequence of inter-dependent random events. Therefore, the events are dependent and are not identically – IID – distributed. In other words events occurring first, affect future failures. The application of, say the Weibull distribution, will therefore result in incorrect results since life data analysis assumes that the events are IID. This statement can be illustrated where the number of repaired failures eventually exceeds the total number of parts in the field, consequently resulting in the failure distribution function being greater than 1.0, which is mathematically impossible.

For many systems in a real world environment, a repair may only be enough to get the system operational again. If the water pump fails on the car, the repair consists only of installing a new water pump. Similarly, if a seal leaks, the seal is replaced but no additional maintenance is done. This is the concept of *minimal repair*. For a system with many failure modes, the repair of a single failure mode does not greatly improve the system reliability from what it was just before the failure. Under minimal repair for a complex system with many failure modes, the system reliability after a repair is the *same-as-old*. In this case, the sequence of failures at the system level follows a non-homogeneous Poisson process (NHPP).

For repairable items which are repaired when they fail, **reliability is the probability that failure will not occur in the period of interest, when more than one failure can occur**. It is characterised by the rate of occurrence of failure (ROCOF) and the mean time between failures (MTBF) under the particular condition of a constant ROCOF. It is often assumed that failures do occur at a constant ROCOF, in which case it is denoted as $\lambda = \frac{1}{\text{MTBF}}$. However, this is only a special case, valuable because it is often true and because it is easy to understand.

If a repairable system can be repaired to an *as-good-as new* condition, then the appropriate model for describing the system's failure behaviour is the *ordinary renewal process* (see O'Connor and Kleyner (2012, 64) for more details). If a system upon repair retains the same failure behaviour as before, referred to as *same-as-old*, it is modelled by the non-homogeneous Poisson process (NHPP). If the condition after repair is however *better-than-old*, but *worse-than-new*, then it is modelled by the *generalised renewal process* (see Kaminskiy and Krivtsov (2000) for more details). (O'Connor and Kleyner 2012, 70, 147; Reliawiki.org 2016).

In conclusion. Sometimes an item may be considered as both repairable and non-repairable. For example, a missile is a repairable system whilst it is in store and subjected to scheduled tests, but it becomes a non-repairable part when it is launched. Reliability analysis of such systems must take account of these separate states. Repairability might also be determined by other considerations. For example, whether an electronic circuit board is treated as a repairable item or not will depend upon the cost of repair. An engine or a vehicle might be treated as repairable only up to a certain age.

3.3 Hazard Rate, Rate of Occurrence of Failure and its Variants

Failure rate, *hazard rate*, *force of mortality (FOM)* and *rate of occurrence of failure (ROCOF)*, are frequently used and have wide meaning in reliability engineering. These concepts are sometimes used interchangeably, which is incorrect.

In general, the terms all refer to the frequency with which parts or systems fail and are applied to both repairable systems and non-repairable parts. In general it can be compared to a car's speedometer. A speedometer indicates how many kilometers per hours the car is travelling at. In reliability engineering it is an indication of how many failures occur per time unit for a specific system or part.

3.3.1 Hazard Rate

From the reliability engineering literature, hazard rate, failure rate and FOM, are all the same. It is a function of the life distribution of a single item and an indication of the "proneness to failure" of the item after a period of time. It is therefore associated with **non-repairable parts** (Rausand and Hoyland, 2004, 9).

In these notes **hazard rate, $\mathbf{h(t)}$**, will be used, instead of the other referenced terminology[2]. Hazard rate is defined as: **the conditional probability of failure in local time (§3.5), given that the failure has not occurred by time, $\mathbf{x}$.** For a part, x is measured from the time it was put into service (Ascher and Feingold, 1984, 156).

There are three basic ways in which the pattern of failures can change with time. The hazard rate may be decreasing, increasing or constant. We can tell much about the causes of failure and about the reliability of the part by appreciating the way the hazard rate behaves in time (O'Connor and Kleyner, 2012, 8).

Decreasing hazard rates are observed in parts which become less likely to fail as their survival time increases. This is often observed in electronic equipment and parts. "Burn-in" of electronic parts is a good example of the way in which knowledge of a decreasing hazard rate is used to generate an improvement in reliability. The parts are operated under failure-provoking stress conditions for a time before delivery. As substandard parts fail and are rejected the hazard rate decreases and the surviving population is more reliable.

A **constant hazard rate** is characteristic of failures which are caused by the application of loads in excess of the design strength, at a constant average rate. For example, overstress failures due to accidental or transient circuit overload, or maintenance-induced failures of mechanical equipment, typically occur randomly and at a generally constant rate.

Wear-out failure modes follow an **increasing hazard rate**. For example, material fatigue brought about

[2] Ascher and Feingold (1984, 136,152) provides an extensive discussion about the misconceptions related to the ambiguous use of the above terminology, which can be consulted.

by strength deterioration due to cyclic loading is a failure mode which does not occur for a finite time, and then exhibits an increasing probability of occurrence.

3.3.2 Rate of Occurrence of Failure (ROCOF)

ROCOF is the hazard rate equivalent for **repairable systems** and is defined **as the unconditional probability of failure in global time (§3.5), t**. For a system, t is measured from the instant at which the system was placed in operation (Ascher and Feingold, 1984, 156). Failure rate is often incorrectly referred to in the context or repairable systems. It should be noted that only ROCOF is attributed to repairable systems and reference to any of the terminology in §3.3.1 is strictly speaking, incorrect.

For repairable systems, there are also three ways in which the pattern of failures can change with time. ROCOF can be decreasing, constant or increasing. For repairable system, times between successive failures tend to increase initially, so that the ROCOF tends to decrease; in the middle region, times between successive failures neither tend to increase nor decrease, and then in the long life region failures of a single system tend to occur more frequently (Ascher and Feingold, 1984, 137).

3.3.3 Bathtub Curve

The combined effect of decreasing, constant and increasing hazard rate and ROCOF is commonly referred to as the bathtub curve (Figure 3.3). However, it is important to realised there are *two* bathtub curves – one for non-repairable parts and one for repairable systems. From a probabilistic perspective the difference between the two is that the hazard rate of a probability distribution is a relative rate, since it is the conditional probability of first and only failure, given survival to the point of failure. To the contrary, the ROCOF of a point process is an absolute rate – the probability that a failure occurs at a point in time. In special cases these rates can be numerically equal but they are obviously not equivalent. The statistical interpretation of the bathtub curve for a repairable system differs greatly from that given for a population of parts, since such systems can fail many times (Ascher and Feingold, 1984, 136).

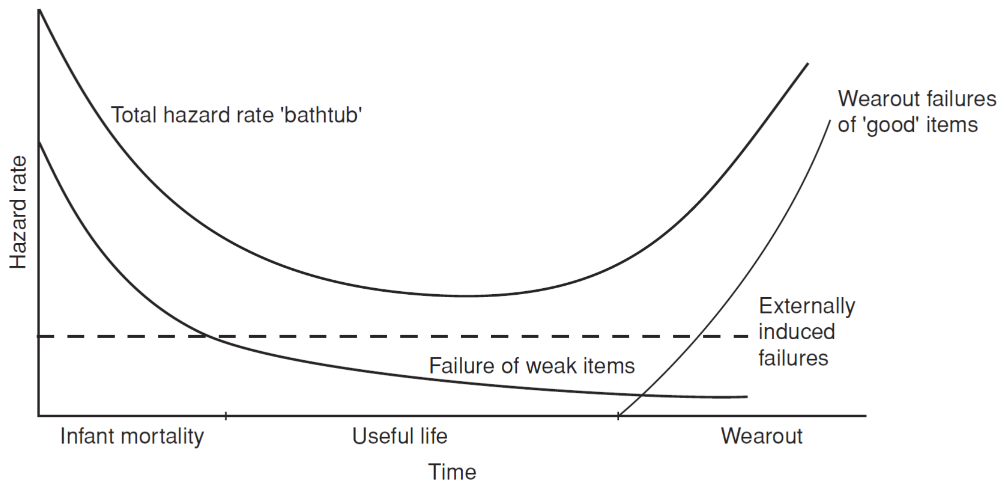
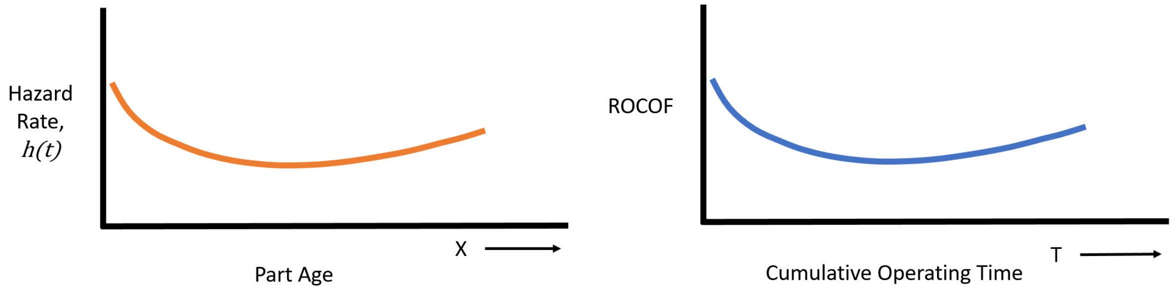


Figure 3.3: The "bathtub" curve (adopted from O'Connor and Kleyner (2012, 9))

Figure 3.4a represents the bathtub curve for non-repairable parts, where the x-axis represents the part age population of a particular life data distribution. Figure 3.4a corresponds to the most bathtub curves illustrated in literature (also Figure 3.3) indicating areas of infant mortality, useful life and wear-out (refer back to §3.3.1). Figure 3.4b represents the bathtub curve of repairable systems, where the x-axis represents the cumulative operating time of one single repairable system over the duration of its life (refer back to §3.3.2).



(a) Bathtub curve for hazard rate, $h(x)$ (b) Bathtub curve for ROCOF

Figure 3.4: Difference between non-repairable parts and repairable systems' bathtub curves (adopted from Ascher and Feingold (1984, 137))

3.4 Effect of Maintenance on Failure Data

According to Vlok (2014), systems are not allowed to age undisturbed in practice. For example, maintenance is performed which should be accounted for when constructing lifetime models. There are three ways to approach the effects of maintenance on failure times:

1. Assume that maintenance on a particular system or a group of similar systems always has the same effect on the lifetime regardless of the frequency or the type of maintenance. Also assume that the effect of maintenance will remain constant for the lifetimes of future systems. This approach essentially ignores the effect of maintenance and would only be followed when information about the occurrences or types of maintenance is unavailable.

2. Include the times of maintenance as truncated failure time observations (also called suspensions) in a data set. Suspensions are considered as partial information by regression coefficients of reliability models. For example it is only known that the system survived up to the age of the maintenance intervention but does not provide any information about when the system would have failed if left undisturbed. For further information on suspensions and censored data refer to O'Connor and Kleyner (2012, 73-75), Turnbull (1976), Lee and Desu (1972), Zhang, Singer, Zhang, and Singer (2010) and Zhang and Singer (2010).

3. Assume that the diagnostic information that describes the condition of a system or the conditions under which a system operates for example vibration levels, temperatures, pressures, load, etc., also quantifies the effects of preventive maintenance. Include these parameters in multivariate lifetime models to better describe a system's recorded and future survival times. This approach is considerably more complex than the first two but also generally yields by far the best results. Publications such as Ma (2008), Ma and 34 Adviser-Krings (2008) or Ma and Krings (2008) are useful introductions to multivariate survival models.

Only the first two approached will be focussed on in this module.

3.5 Failure Data and Timelines

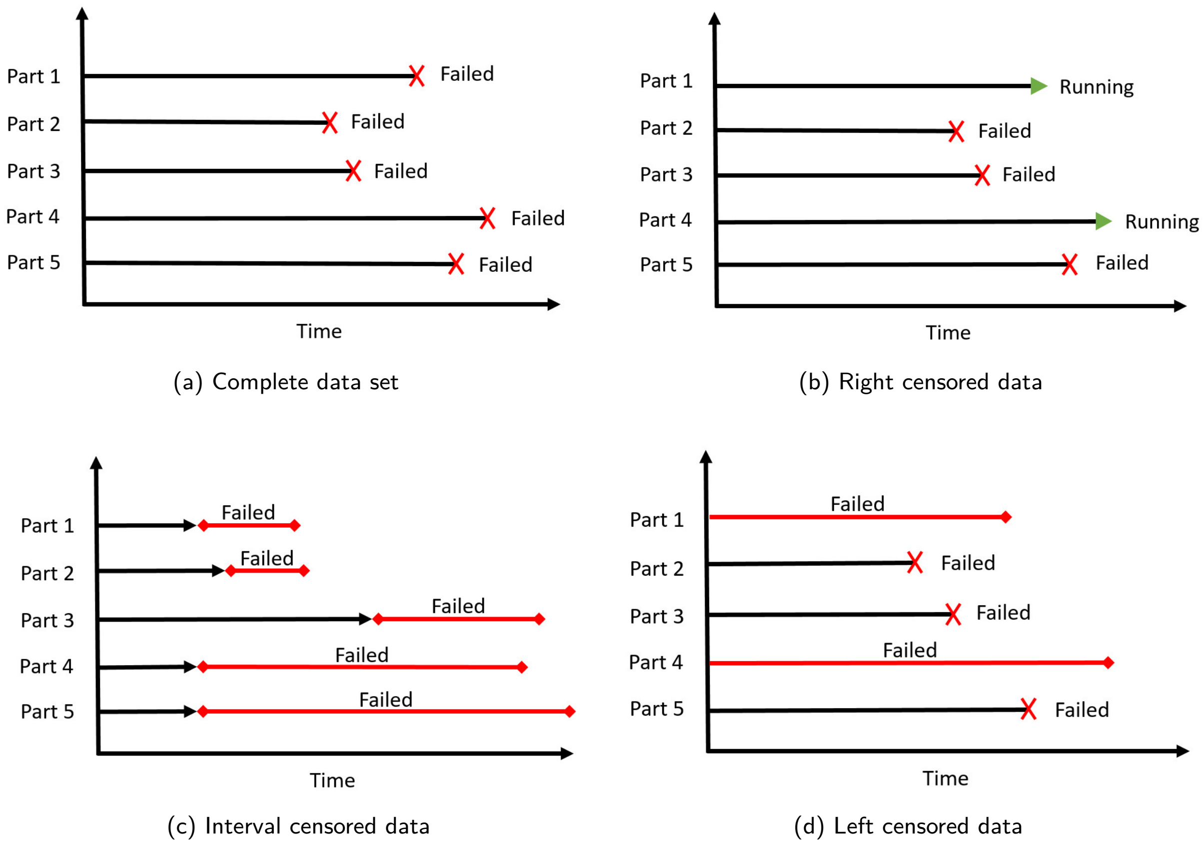
The acquisition of failure data and understanding of failure timelines are essential before embarking on reliability data analysis.

3.5.1 Data Acquisition

Data required for lifetime failure analysis is often difficult to obtain in practice. The best source of failure history data are from EAMS or CMMS which are widely used in industry. However, often the ineffective use of these systems result in insufficient data or data quality for analysis purposes. Another source of failure history is to review process data (which is often recorded at rates as high as 1 Hz) from where failure times can be derived. Other sources for information include shift logbooks and financial records. The quality of data is however important. The usefulness of reliability models are directly proportional to the quality of the data that it is based on.

A further assumption that needs to be satisfied is that the collected data are representative of the population of interest. If this is not the case, the results will almost always be bad estimations. In life data analysis, ideally all available data should be used, but sometimes data are incomplete.

Complete data means that the value of each sample part is known. This means that the data set would consists of the times-to-failure of all parts in the sample. For example, if five parts where tested and failed, with times-to-failure recorded, the data would be complete (Figure 3.5a). In many cases data is not complete



(a) Complete data set

(b) Right censored data

(c) Interval censored data

(d) Left censored data

Figure 3.5: Different life data distribution classifications

or the times-to-failure of all the parts are not known. This is called *censored data*. The most common case of sensored data is *right sensored data* or *suspended data*. These data sets includes parts which have not failed. For example five parts are tested, but only three failed by the end of the test (Figure 3.5b). The two parts which did not fail are right censored or suspended data. The term *right* means that the event of interest is to the right of the data point, meaning that if the test continued the failure would eventually occur.

Interval censored data is associated with uncertainty of the exact times when the parts failed within an interval (Figure 3.5c). This type of data is associated with tests or situations where the parts are not constantly monitored. It is recommended to avoid interval censored data, since it is less informative than completed data.

Left censored data is similar to interval data. In left censored data a failure is only known to be before a certain time (Figure 3.5d).

3.5.2 Failure Timelines

An understanding of the time scales for measuring lifetimes is required, before lifetime models for non-repairable parts and repairable systems can be introduced. Figure 3.6 represents a system (non-repairable or repairable) which has failed along a sample path. The first two instants in figure 3.6 represent real failures, and the last a suspension. Further explanation is according to the work of Ascher and Feingold (1984, 18-19).

In figure 3.6, X_i, represents the *interarrival* time between the $(i-1)^{th}$ failure and the i^{th} failure, where X_i is a random variable, $i = 1$, 2, 3,... and with $X_0 \equiv 0$. The real variable x_i measures the time elapsed since the most recent failure. Interarrival time is referred to as *local time* an is used when analysing non-repairable parts.

To account for right censored failure observations an auxiliary variable C_i is recorded for each X_i for indicating whether X_i represents an actual failure or a censored failure (suspension). For a suspension, $C_i = 0$, while for an actual failure $C_i = 1$).

In figure 3.6, T_i measures total time from $t = 0$ to the i_{th} failure and is referred to as the *arrival* time to the i^{th} failure. This time scale is referred to as *global time* and is used when analysing repairable systems. It follows that $T_r = X_1 + X_2 + \cdots + X_r$, where r is the total number of observed failures. Similar to interarrival times, C_i is also recorded for T_i.

The real variable $N(t)$ represents the maximum value of r for which $T_r \leq t$. For example $N(t)$ is the number of failures which occur during $(0, t]$. $N(t), t \geq 0$ is the integer value counting process which includes both the number of failures in $(0, t]$, $N(t)$, and the instants $T_1, T_2 \ldots$.

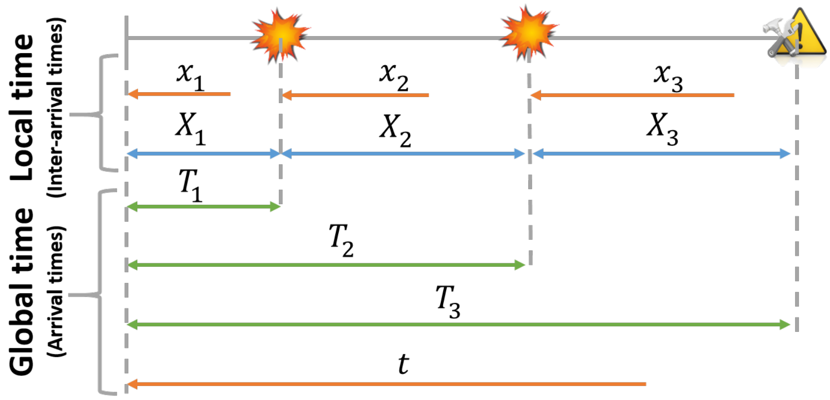


Figure 3.6: A sample path of a failure process (Adapted from Ascher and Feingold (1984, 18))

3.6 Model Selection

A failure data set needs to be assessed for certain characteristics to determine which lifetime model is most appropriate for analysis. Figure 3.7 illustrates the model selection process which is based on the work by Ascher and Feingold (1984, 72). The next sections describe the particulars of each step in the process.

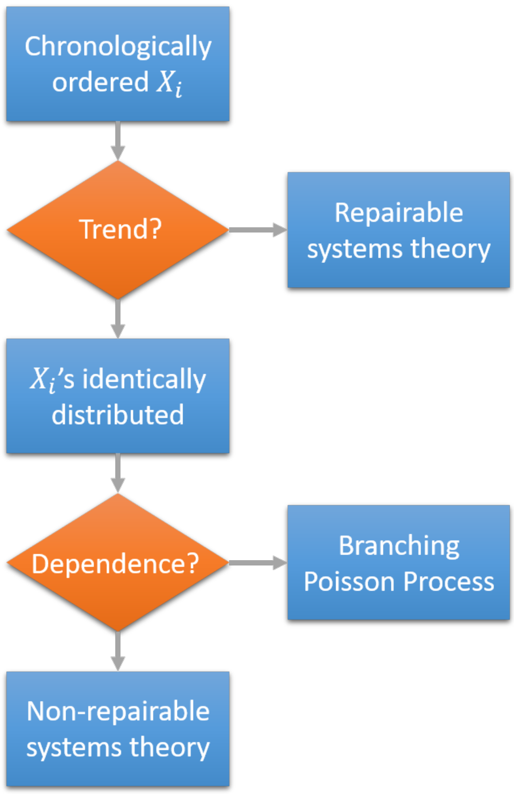


Figure 3.7: Process for selecting the appropriate lifetime model for reliability analysis (Adapted from Ascher and Feingold (1984, 72))

3.6.1 Events Table with Chronologically Ordered Data

The following steps are proposed to avoid unnecessary calculation errors:

1. The process starts by creating an events table with *chronologically* ordered interarrival times, X_i's. It is important that the data is in chronological order and has not been ranked according to magnitude. Generate an *events table* which contains the interarrival (local) times, X_i, and arrival (global) times, T_i of the system to be analysed and include the auxiliary variable C_i (refer to §3.5) for each failure observation

2. Decide on an appropriate time scale, i.e. local time or global time, and apply it consistently. Local time can be converted to global time and vice versa. Remember local time is use for non-repairable parts and global time for repairable systems (refer to §3.5).

i	X_i [days]	T_i [days]	C_i
1	129	129	1
2	161	290	1
3	151	441	1
4	57	498	1
5	111	609	1
6	127	736	1
7	131	867	0

Table 3.5: Example events table

Table 3.5 shows an example events table. Note that failure observation, $i = 7$ is a suspension, as indicate by C_7. The proper structuring of the data set is important as most calculation errors are a result of poorly structured data sets.

Chronological X_i's are tested for a trend – a tendency to increase corresponds to improvement, and decrease to deterioration. The existence of a trend shows that the data are non-stationary, which mean that even the marginal distributions of the X_i's are not identical – in other words not IID. It would therefore be meaningless to investigate the properties of a non-existent common distribution of the X_i's if a trend is present (Ascher and Feingold, 1984, 71). For this reason it is important to use chronologically ordered data, since ranking data according to magnitude – which implies IID – will give misleading results. Whenever failure data are reordered all trend information is lost (O'Connor and Kleyner, 2012, 341).

***Example* 3.3.** *Table 3.6 consists of interarrival times (colum 1), chronologically ordered arrival times (colum 2) and data in rank order (IID) (colum 3). A trend test of this data reveals in increasing trend, meaning that the interarrival times are not IID. If however the data was assume to be IID (column 3) and plotted on probability paper the resulting plot (figure 3.8) would be Line A, which is apparently exponential component life distribution. This is obviously incorrect, as shown by Line B which represents the chronological data in colum 2 (O'Connor and Kleyner, 2012, 340). It should be noted that the lack of a trend is not absolute proof of stationarity, although the possiblity is usually ignored in practice (Cox and Lewis 1966, 61; Ascher and Feingold 1984, 71).*

1	***2***	***3***
$\mathbf{X_i}$	***Chronological*** $\mathbf{T_i}$	***Ranked*** $\mathbf{X_i}$
175	*175*	*12*
21	*196*	*14*
108	*304*	*21*
111	*415*	*23*
89	*504*	*38*
12	*516*	*47*
102	*618*	*51*
23	*641*	*89*
38	*679*	*102*
47	*726*	*108*
14	*740*	*111*
51	*791*	*175*

Table 3.6: *Comparison between chronological and ranked failure data*

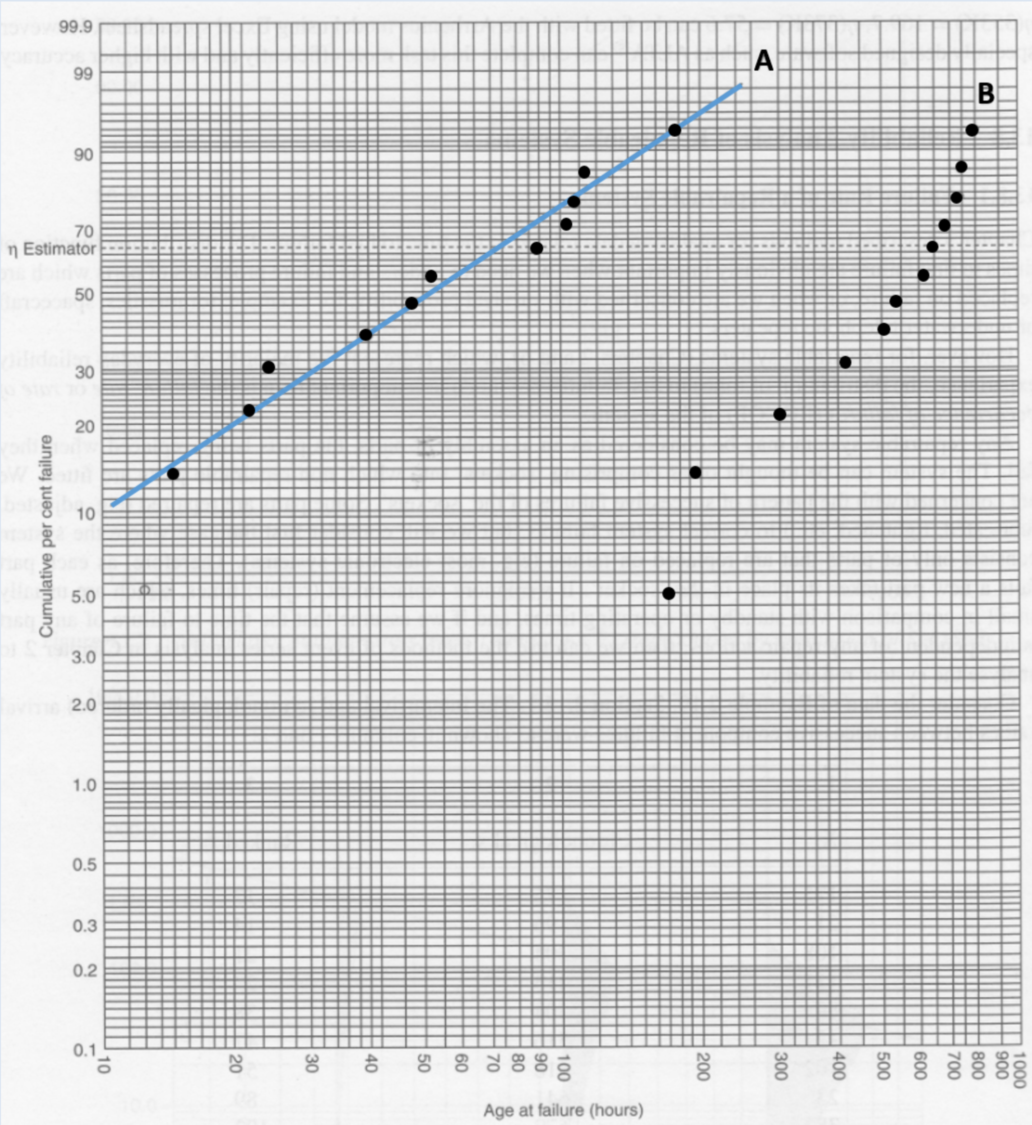


Figure 3.8: *Plot of chronological and ranked data on Weibull paper*

3.6.2 Laplace's trend test

The Laplace trend test is a simple and powerful test for distinguishing between a constant rate at which events are occurring and an increasing or decreasing rate of occurrence of such event (Ascher and Hansen, 1998, 453).

Laplace makes use of the fact that under the homogenous Poisson process (HPP) assumption, the first $r-1$ arrival times, T_1, T_2, ..., T_{r-1} are the order statistics from a uniform distribution on $(0,T_r]$ (Ascher and Feingold, 1984, 78-79). Therefore,

$$U = \frac{\frac{\sum_{i=1}^{r-1} T_i}{r-1} - \frac{T_r}{2}}{T_r\sqrt{\frac{1}{12(r-1)}}} \tag{3.8}$$

The Laplace trend test compares the centroid of the observed arrival times with the mid-point of the period of observation. If $U = 0$ there is no trend, which means the process is stationary. If $U \leq 0$ the trend is decreasing indicating the interarrival times are tending to become larger, while for $U \geq 0$ the trend is increasing and interarrival times are tending to become progressively smaller (O'Connor and Kleyner 2012, 63; Ascher and Feingold 1984, 71).

U approximates a standardised normal variate with approximation being adequate at a 5% level of significance for $r \geq 4$. The associated critical value for this level of significance is, $z = 1.96$. If $-1.96 \geq U \geq 1.96$, then there is not sufficient evidence to reject the hypothesis (H_0) of no trend. In the case where $U \geq 1.96$ there is however strong evidence of reliability degradation while $U \leq -1.96$ indicates reliability improvement. For simplicity the rejection criteria is approximated at $U \geq 2$ and $U \leq -2$ (O'Connor and Kleyner 2012, 63; Ascher and Feingold 1984, 71).

For cases where $2 > U > 1$ or $-1 > U > -2$, Laplace's trend test is not able to indicate with certainty whether a trend is present in the data set or not. Alternative tests such as Lewis and Robinson (1974), Mann (1945) and Anderson and Darling (1954) are used in these cases. Figure 3.9 summarises the Laplace trend test outcomes.

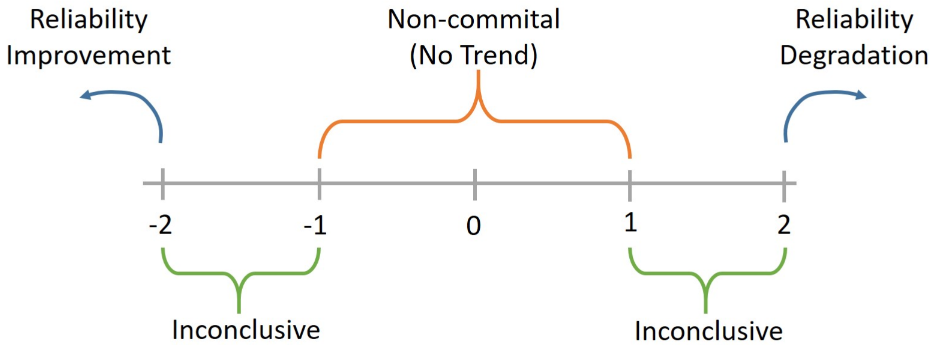


Figure 3.9: Graphical explanation of Laplace's trend test outcomes

***Example* 3.4.** *The following dataset consists of interarrival times between failures (in months) for a pump:*

64, 43, 17, 21, 94, 48, 3, 13, 96, 91, 16, 63, 38, 69, 50

Equation 3.8 is defined in global time and therefore the interarrival times (X_i's) need to be converted to arrival times (T_i's):

64, 107, 124, 145, 239, 287, 290, 303, 399, 490, 506, 569, 607, 676, 726

Substitution into (3.8) produces,

$$\begin{aligned} U &= \frac{\frac{4806}{14} - \frac{726}{2}}{726\sqrt{\frac{1}{12(15-1)}}} \\ &= -0.35196 \end{aligned} \tag{3.9}$$

The result above shows that no trend is present in the data set.

3.6.3 Test for Dependency

If there is no proof of a trend it means that successive X_i's have identical marginal distributions, but it is not necessarily independent. According to Cox and Lewis (1966) the test for dependence is normally omitted since approximately 30 failure observations are required to perform a test for dependence with reasonable confidence. Since it is rare to have more than 30 failure observations available for modelling independence it is normally assumed, which means the data is IID and can be modelled as a non-repairable part according to life data analysis (refer to §3.2.1). It is however important to know that a positive test for dependence would mean a model such as the Branching Poisson Process will need to be used. Refer to Cohen (1988) and Gretton et al. (2007) for further details about dependence.

3.7 Analysing Non-Repairable Systems

When successive interarrival times (X_i's) of the dataset do not exhibit trend or dependence the data is adequately represented by a renewal process. In other words there is no indication that the data is not IID and life data analysis (§3.2.1) can be performed (Ascher and Feingold, 1984, 91). From here on neither the order of data or the particular machine on which the event occurred are important any more.

The goal of life data analysis is to find the best fitting statistical distribution and to derive estimates of the distributions parameters to consequently describe reliability related functions. The Weibull distribution is popular for analysing life data, and so the process is often referred to as *Weibull analysis*. There are however other life distributions which can also be used. These include the exponential, extreme value, lognormal and normal distributions (O'Connor and Kleyner, 2012, 71). Refer to O'Connor and Kleyner (2012, 33-51) for further information on the use of these distributions.

In the life data analysis the analyst chooses the life distribution which is the most appropriate to model a particular dataset based on goodness-of-fit tests, experience and engineering judgement. In this course however only the Weibull distribution will be covered for analysing non-repairable parts.

3.7.1 Variable Declaration and Assumptions

a Time required to perform preventive maintenance

b Time required to perform maintenance on a non-repairable part that failed unexpectedly

$R_X(x)$ Probability of system survival up to instant x

$F_X(x)$ Probability density function or Probability of system failure before instant x

$h_X(x)$ Hazard rate

$L(X,\theta)$ Likelihood function for the probability density

β Shape parameter for the Weibull distribution

η Scale parameter for the Weibull distribution

θ Vector of unknown parameters in L

E Expected value

μ_{r+1} Residual life

$\underset{\sim}{X}_{r+1}$ Lower confidence limit of the residual life estimate

$\tilde{X}_{r+1}$ Upper confidence limit of the residual life estimate

X_p Recommended preventive replacement time

X_c Total cycle time of a non-repairable part

C_{X_p} Total cycle cost if the non-repairable part is always maintained at time X_p or at failure if $X_{r+1} < X_p$

$C(X_p)$ Total cycle cost per unit time as a function of X_p

w Number of system

C_p Cost of preventive maintenance

C_f Cost of unexpected failure

m Total number of observed failures

x Continuous time

X Discrete event time measured in local time

r Total number of observed events

T Discrete event time measured in global time

U Laplace

3.7.2 Weibull Distribution

Weibull probability data analysis is the most widely used in analysing failure data. It is popular because by adjusting the distribution parameters it can be made to fit many other distributions. It's advantages therefore are: its flexibility, easy interpretation, relation to failure rates and the bathtub curve (O'Connor and Kleyner, 2012, 37, 78). The distribution is given by,

$$f_X(\mathsf{x}) = \frac{\beta}{\eta}\left(\frac{\mathsf{x}}{\eta}\right)^{\beta-1} \cdot e^{-\left(\frac{x}{\eta}\right)^{\beta}} \tag{3.10}$$

where β is the shape and η the scale parameter of the distribution. η is also known as the *characteristic of life* which is the life at which 63.2% of the population will have failed (O'Connor and Kleyner, 2012, 38). $f_X(x)$ provides the probability of system failure at instant x.

If $f_X(x)$ is integrated with respected to x, the probability of system failure before a certain instant, x, is obtained,

$$F_X(x) = 1 - e^{-\left(\frac{x}{\eta}\right)^\beta} \tag{3.11}$$

where $F_X(\infty) = 1$. From F_X(x) the probability of system survival up to a certain instant x, can be derived. This reliability function, R_X(x) is given by,

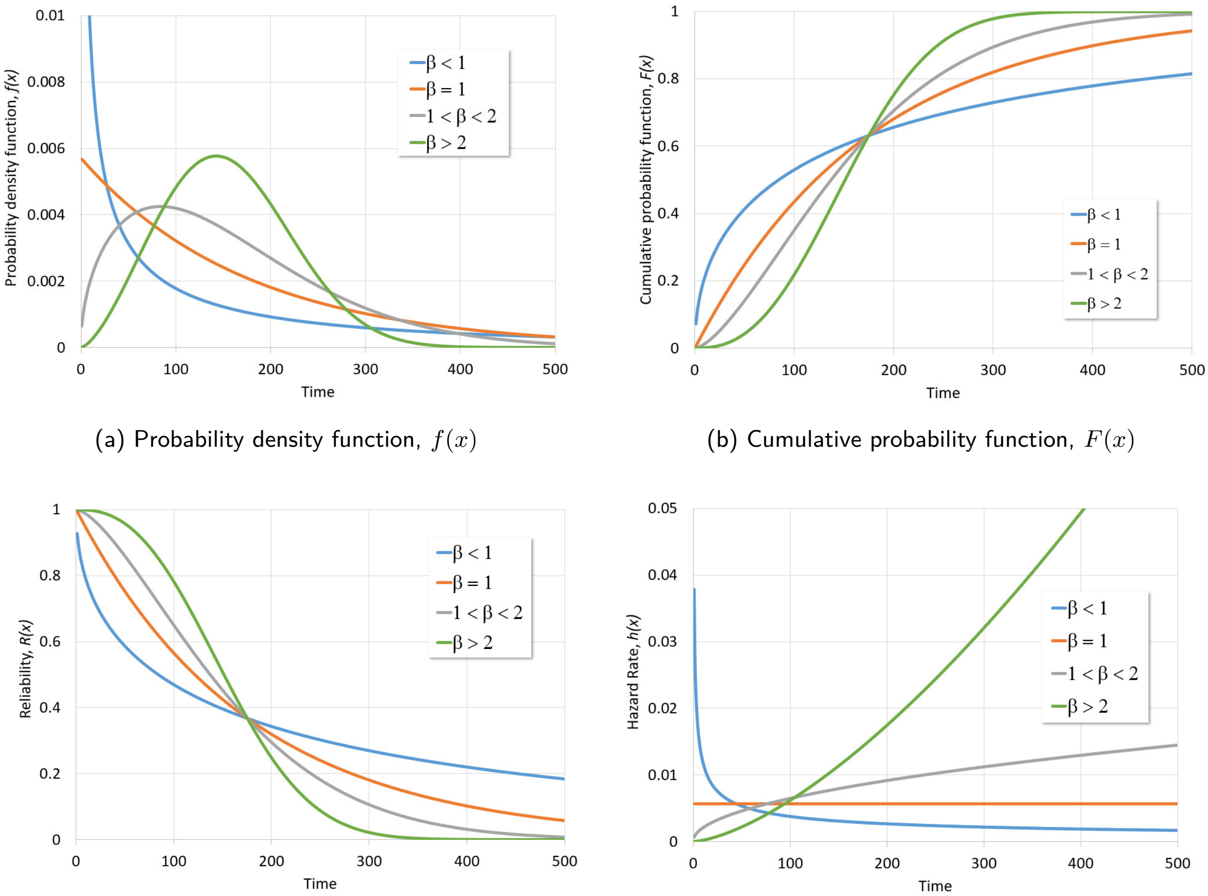
$$R_X(x) = e^{-\left(\frac{x}{\eta}\right)^\beta} \tag{3.12}$$

where $R_X(\infty) = 0$. The ratio of $f_X(x)$:$R_X(x)$ yields a conditional probability referred to as the *hazard rate* [3] of a non-repairable part:

$$h_X(x) = \frac{f_X(x)}{R_X(x)} \tag{3.13}$$

$$= \frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1} \tag{3.14}$$

Figure 3.10 illustrates the Weibull functions for different values of β.



(a) Probability density function, $f(x)$

(b) Cumulative probability function, $F(x)$

(c) Reliability, $R(x)$

(d) Hazard rate, $h(x)$

Figure 3.10: Reliability functions for different values of β

[3]Also referred to as Force of Mortality (FOM) or conditional intensity.

3.7.3 Bathtub Curve Relationships

The β-parameter reflects the hazard function or expected failure rate of the Weibull distribution. Conclusions can be made about the failure characteristics from the dataset by considering β in relation to unity. The combined effect of different hazard rates produces the bathtub curve in figure 3.11. The following guide assists in interpreting the results based on the value of β (O'Connor and Kleyner, 2012, 84):

$\boldsymbol{\beta > 1}$ For a hazard rate increasing with x, i.e. $\beta > 1$ (see figure 3.10d), a non-repairable part is in a *wear-out* phase and has an increasing probability to fail. Usage-based maintenance should be considered under such circumstances, subject to the assessment of safety and cost considerations. Preventive action will only be feasible if the total cost of failure is considerably higher than the total cost of prevention.

$\boldsymbol{\beta \approx 1}$ If (3.13) remains constant with x, i.e. $\beta \approx 1$, the non-repairable part has a random failure pattern because the instantaneous risk of failure remains constant throughout its life. Run-to-failure maintenance should be considered if condition monitoring techniques are not feasible.

$\boldsymbol{\beta < 1}$ For n non-repairable part with a decreasing hazard rate, where $\beta < 1$, run-to-failure should be considered, since the probability of failure reduces as time progresses. Condition-based maintenance (e.g. vibration analysis, thermography) could be used for any type of the hazard rate, provided that it is safe and technically and financially viable.

$\boldsymbol{\beta > 6}$ It is time to become suspicious. Although $\beta > 6$ is not uncommon it reflects accelerated rate of failure and fast wearout, which is more common in brittle parts (consult O'Connor and Kleyner (2012, 84) for further interpretation of high β values).

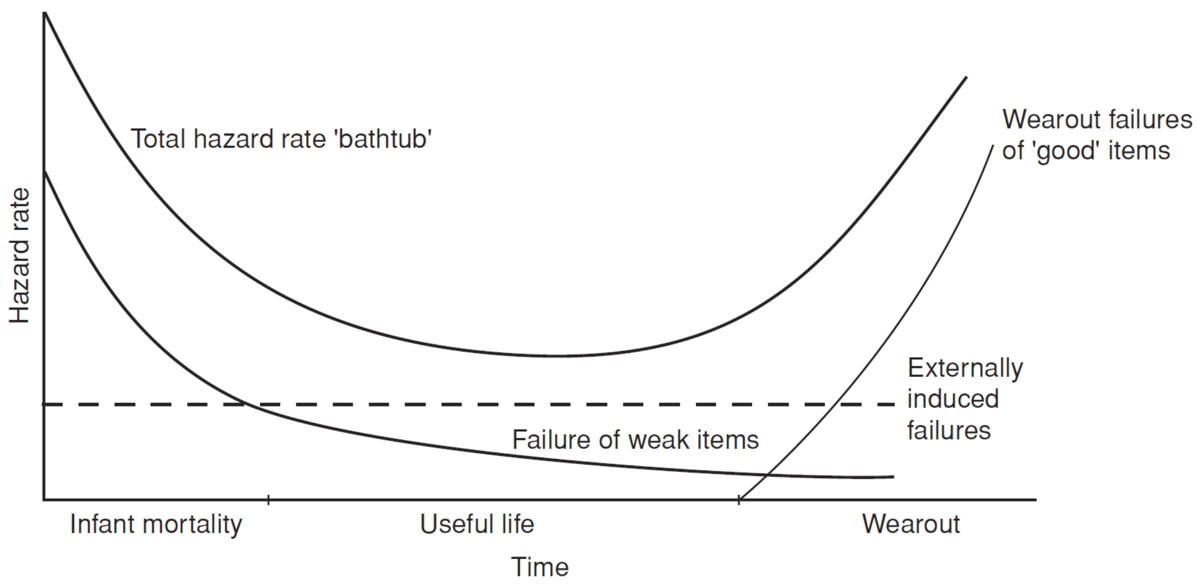


Figure 3.11: The combined effect of different hazard rates produces the so-called *bathtub curve* (Adapted from O'Connor and Kleyner (2012, 9))

3.7.4 Parameter Estimation

Parameter estimation for the Weibull distribution is reasonably easy and can be obtained numerically by maximizing the likelihood given by,

$$lnL(X,\theta) = \sum_{i=1}^{m}\left[ln\frac{\beta}{\eta} + (\beta-1)ln\frac{X_i}{\eta}\right] - \sum_{j=1}^{r}\left(\frac{X_j}{\eta}\right)^{\beta} \tag{3.15}$$

which produces estimated values for β and η. Care should be taken to distinguish between observed failure times and observed suspensions. Refer to Wasserman (2003),

3.7.5 Residual Life Estimation

The underlying failure process is described by the Weibull distribution in §3.7.2. It is possible to predict the arrival time of the next event. Suppose a system has been in operation for x time units and a maintenance policy exists where the system is replaced preventively at time X_p or at failure, whichever comes first. The conditional expectation of X_{r+1}, (where $X_{r+1} \leq X_p$), is given by (Coetzee, 2004, 405),

$$E[X_{r+1}|X_{r+1} \leq X_p] = \frac{\int_x^{X_p} x \cdot f_X(x)dx}{\int_x^{X_p} f_X(x)dx} \tag{3.16}$$

Therefore, the residual life to the $(r+1)^{th}$ failure is expected to be,

$$\mu_{r+1} = E[X_{r+1}|X_{r+1} \leq X_p] - x \tag{3.17}$$

It is also possible to calculate confidence levels for the residual life estimate. The lower limit, $\underset{\sim}{X}_{r+1} - x$, can be obtained by solving numerically for $\underset{\sim}{X}_{r+1}$ in,

$$\frac{\int_x^{\underset{\sim}{X}_{r+1}} f_X(x)dx}{1 - \int_0^x f_X(x)dx - \int_{X_p}^{\infty} f_X(x)dx} = 0.025 \tag{3.18}$$

Similarly is the upper limit $\tilde{X}_{r+1} - x$ can be obtained by solving for $\tilde{X}_{r+1}$ in,

$$\frac{\int_x^{\tilde{X}_{r+1}} f_X(x)dx}{1 - \int_0^x f_X(x)dx - \int_{X_p}^{\infty} f_X(x)dx} = 0.975 \tag{3.19}$$

In the case where there is no preventive maintenance intervention, in other words where $X_p = \infty$ and the residual life is required shortly after X_r, $x \approx 0$, and (3.16) becomes,

$$E[X_{r+1}] = \frac{\int_0^{\infty} x \cdot f_X(x)dx}{\int_0^{\infty} f_X(x)dx} \tag{3.20}$$

The corresponding residual life to the $(r+1)^{th}$ failure is expected to be,

$$\mu_{r+1} = E[X_{r+1}] - x \tag{3.21}$$

The confidence limits reduce to:

$$\int_0^{\underset{\sim}{X}_{r+1}} f_X(x)dx = 0.025 \tag{3.22}$$

and :

$$\int_0^{\tilde{X}_{r+1}} f_X(x)dx = 0.975 \tag{3.23}$$

This is a simplified approach for calculating residual life, which results in a broad confidence band. More advanced techniques are available for predicting more accurate confidence limit.

***Example* 3.5.** *The data from Example 3.4 is presented below together with the auxiliary variable, C_i given in brackets. $C_i = 1$ denotes a real failure and $C_i = 0$ represents a suspension or preventive replacement:*
64 (1), 43 (1), 17 (0), 21 (1), 94 (1), 48 (0), 3 (1), 13 (1), 96 (0), 91 (1), 16 (1), 63 (0), 38 (1), 69 (1), 50 (1)
Since it is known from Example 3.4 that this data is non-committal and associated with a non-repairable part, the Weibull distribution is fitted according to the maximum likelihood method (3.15), which produces the following probability density function,

$$f_X(x) = \frac{1.404}{65.102}\left(\frac{x}{65.102}\right)^{1.404-1} e^{-\left(\frac{x}{65.102}\right)^{1.404}}$$

Suppose it is of interest what the probability of the system's success up to 40 days is. It is then required to calculate $R_X(40)$ by using (3.12),

$$\begin{aligned} R_X(40) &= e^{-\left(\frac{40}{65.102}\right)^{1.404}} \\ &= 60.38\% \end{aligned}$$

If the current system is $x = 20$ days old and a preventive replacement intervention exists at $X_p = 80$, the next expected failure time, X_{r+1} according to (3.16) is,

$$\begin{aligned} E[X_{r+1}|X_{r+1} \le 80] &= \frac{\int_{20}^{80} x \cdot f_X(x)dx}{\int_{20}^{80} x \cdot f_X(x)dx} \\ &= 47.46 \text{ days} \end{aligned}$$

The residual life of the current system is therefore $\mu_{r+1} = 47.46 - 20 = 27.46$ days. Equations (3.18) and (3.19) is used to calculate a confidence bounds of 95% around the estimate, which results in,

$$\begin{aligned} \underset{\sim}{X}_{r+1} - 20 &= 1 \text{ days} \\ \tilde{X}_{r+1} - 20 &= 57 \text{ days} \end{aligned}$$

3.7.6 Long Term Cost Optimisation

Together with residual life estimation it is often required to find an appropriate preventive maintenance instance, X_p, which will minimise the maintenance cost of the non-repairable part over it life. The total cost relating to an unexpected failure can be denoted by C_f, while the total cost related to preventive maintenance is C_p. The total cost of maintaining the system when considering it is always replaced at failure or at X_p is (Elsayed 2012, 502; Jardine and Tsang 2006, 50),

$$C_{X_p} = C_p \cdot R(X_p) + C_f \cdot F(X_p) \tag{3.24}$$

where it is assumed that some unexpected failures will occur despite the preventive maintenance intervention. The expected life cycle time of the non-repairable part at time $x = 0$ is given by,

$$E[X_c] = (X_p + a) \cdot R(X_p) + (E[X_{r+1}|X_{r+1} \leq X_p] + b) \cdot F(X_p) \tag{3.25}$$

where a represents the time required for a preventive replacement and b the time necessary for a failure replacement. It follows that the cost per unit time is,

$$C[X_p] = \frac{C_p \cdot R(X_p) + C_f \cdot F(X_p)}{(X_p + a) \cdot R(X_p) + (E[X_{r+1}|X_{r+1} \leq X_p] + b) \cdot F(X_p)} \tag{3.26}$$

The minimum value of $C(X_p)$ is obtained by differentiating (3.26) with respect to X_p or by graphical inspection.

***Example* 3.6.** *For the data of Example 3.5, assume that the total cost of an unexpected failure, is $C_f =$ R6 000, while the total cost of a preventive replacement is $C_p =$ R1 000. Take $a = 2$ hours $= 0.0833$ days and $b = 8$ hours $= 0.333$ days. The result of substituting these values in (3.26) are, shown in figure 3.12. An analytical optimum exists at 42 days where $C(42) =$ R89.76 per day.*

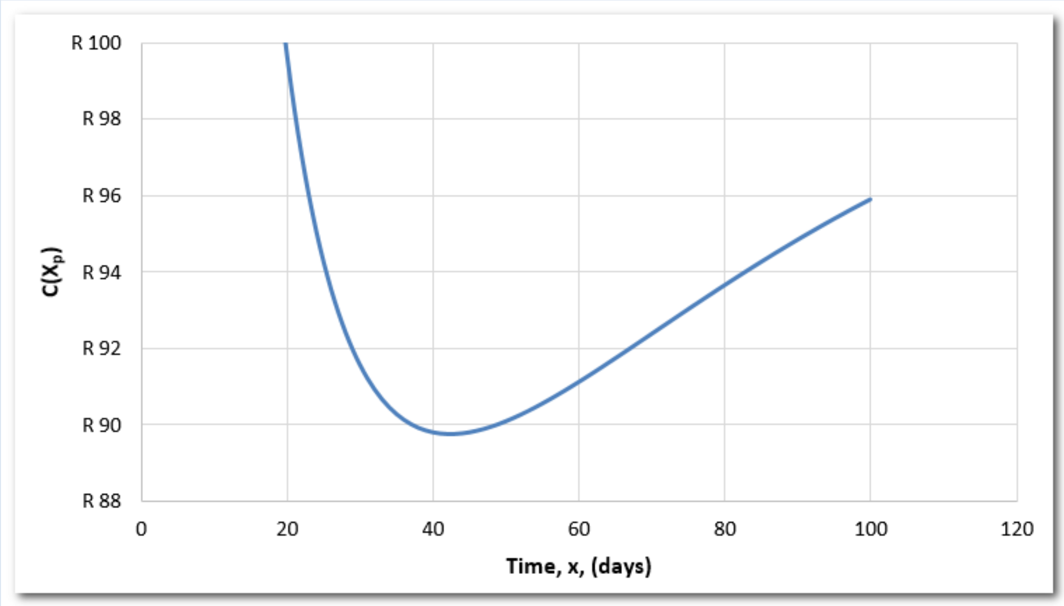


Figure 3.12: *Long term cost optimistation plot with minimum at 42 days and R89.76 per day.*

***Example* 3.7.** *A Weibull distribution is fitted to failure data (with no underlying trend) for an electrical motor and the parameters are estimated as: β = 2.3 and η= 150 days. It costs R1 000 to perform a preventive replacement on the pump and R7 500 to replace the pump in case of an unexpected failure. Two hours are required to perform the preventive replacement and 8 hours are required to replace the pump after failure. By using this information calculate the following:*

1. *The optimal preventive maintenance instant, X_p;*
2. *The long term daily cost if preventive action is always taken at X_p or failure, whichever comes first;*
3. *Assume a newly installed pump is 30 days old and still in operation. Calculate the expected residual life, given X_p. Give your answer within a 90% confidence band.*

Firstly, the optimal preventive maintenance instant X_p is found where the long term cost C is a minimum according to (3.26):

$$C[X_p] = \frac{1000 \cdot R(X_p) + 7500 \cdot F(X_p)}{(X_p + 2/24) \cdot R(X_p) + (E[X_{r+1}|X_{r+1} \leq X_p] + 8/24) \cdot F(X_p)} \tag{3.27}$$

where,

$$E[X_{r+1}|X_{r+1} \leq X_p] = \frac{\int_{x=0}^{X_p} x \cdot f_X(x)dx}{\int_{x=0}^{X_p} f_X(x)dx} \tag{3.28}$$

Solving the equations numerically results in $X_p = 60$ days and $C(X_p) =$ R 30.01 per day.

Secondly, the residual life of a pump that is currently 30 days old, given $X_p = 60$, is determined by using (3.16) and (3.17),

$$E[X_{r+1}|X_{r+1} \leq X_p] = \frac{\int_{x=30}^{60} x \cdot f_X(x)dx}{\int_{x=30}^{60} f_X(x)dx} = 47.38 \textit{ days} \tag{3.29}$$

which results in a residual life of 17.38 days. The confidence limits are obtained through (3.18) and (3.19) which is $\underset{\sim}{X}_{r+1} - x = 2$ *days and* $\tilde{X}_{r+1} - x = 28$ *days.*

3.8 Analysing Repairable Systems

In the previous section, the value of the conditional intensity for non-repairable part, i.e. the FOM, became clear. Ideally we would also prefer to model the conditional intensity for repairable systems but it implies a series of truncated FOMs which is impractical because of data constraints. The *mean intensity* for repairable systems, i.e. the Rate of Occurrence of Failure (ROCOF) or the time derivative of the expected number of failures up to a certain instant, will have to suffice in this case. The Non-homogeneous Poisson Process (NHPP) has been proven as a good representation of repairable systems' ROCOF. Two formats of the NHPP are common in reliability and are discussed in sections to follow (Ascher and Feingold, 1984, 83).

3.8.1 Variable declaration and assumptions

- t Continuous global time
- T Discrete global time
- N($t_1 \to t_2$) Expected number of events from t_1 to t_2
- R($t_1 \to t_2$) Reliability of a repairable system from t_1 to t_2
- $\rho_1(t)$ Log-linear NHPP
- $\rho_2(t)$ Power Law NHPP
- α_1, α_2 Parameters required for the log-linear NHPP
- δ, λ Parameters required for the power law NHPP
- r Total number of observed events
- μ_{r+1} Residual life estimate
- $MTBF_{\rho_1}$ Mean time between failures estimated by $\rho_1(t)$
- $MTBF_{\rho_2}$ Mean time between failures estimated by $\rho_2(t)$
- l_{ρ_1} Log liklihood of $\rho_1(t)$
- l_{ρ_2} log likelihood of $\rho_2(t)$
- C_m Average cost of minimal repair to a repairable system, i.e. restore it to a "best-as-old" (BAO) condition
- C_s Average cost of complete system replacement
- C_R Maintenance cost per unit time of a repairable system
- T* Optimal system replacement instant in terms of time
- N* Optimal system replacement instant in terms of N(0 $\to$ t)

It is assumed that all events are categorised as failures except for the last observation that may be a failure or suspension for the sake of simplicity.

Log-linear NHPP

The log-linear NHPP is given by,

$$\rho_1(t) = exp(\alpha_0 + \alpha_1 t) \tag{3.30}$$

where $\alpha_0 > 0$ for repairable systems. By definition the expected number of failures, N, between any two instants, t_2 and t_1, in the process can be obtained by integrating equation (3.30), i.e.

$$E[N(t_1 \to t_2)] = \frac{1}{\alpha_1}[exp(\alpha_0 + \alpha_1 t_2) - exp(\alpha_0 + \alpha_1 t_1)] \tag{3.31}$$

Therefore it is possible to estimate the system's MTBF from t_1 to t_2 by,

$$MTBF_{\rho_1}(t_1 \to t_2) = \frac{\alpha_1(t_2 - t_1)}{exp(\alpha_0 + \alpha_1 t_2) - exp(\alpha_0 + \alpha_1 t_1)} \tag{3.32}$$

In a similar manner, the reliability of the repairable system from t_1 to t_2 can be calculated by,

$$R(t_1 \to t_2) = e^{-[exp(\alpha_0 + \alpha_1 t_2) - exp(\alpha_0 + \alpha_1 t_1)]/\alpha_1} \tag{3.33}$$

Parameter estimation for repairable systems is easier than for non-repairable part. One possibility is by the simple least-squares parameter estimation method (described in most textbooks of statistics). The difference between the observed number of failures and the number of failures expected by $\rho_1(t)$ should be minimized by,

$$min(\hat{\alpha}_0, \hat{\alpha}_1) : \sum_{i=1}^{r}[E[N(0 \to T_i)] - N(0 \to T_i)]^2 \tag{3.34}$$

for the case where T_r is a failure and

$$min(\hat{\alpha}_0, \hat{\alpha}_1) : \sum_{i=1}^{r-1}[E[N(0 \to T_i)] - N(0 \to T_i)]^2 \tag{3.35}$$

where T_r is a suspension.

Another possibility is to use the Maximum Likelihood method, i.e. maximizing the log likelihood, $I_{\rho_1}(\alpha_0, \alpha_1)$, given by,

$$I_{\rho_1}(\alpha_0, \alpha_1) = r\alpha_0 + \alpha_1 \sum_{i=1}^{r} T_i - \frac{e^{\alpha_0}(e^{\alpha_1 T_r} - 1)}{\alpha_1} \tag{3.36}$$

such that,

$$max(\hat{\alpha}_0, \hat{\alpha}_1) : I_{\rho_1}(\alpha_0, \alpha_1) = I_{\rho_1}(\hat{\alpha}_0, \hat{\alpha}_1) \tag{3.37}$$

for the case where T_r is a failure. If T_r is a suspension, the likelihood function becomes,

$$I_{\rho_1}(\alpha_0, \alpha_1) = r\alpha_0 + \alpha_1 \sum_{i=1}^{r} T_i - \frac{e^{\alpha_0}(e^{\alpha_1 T_r} - 1)}{\alpha_1} - \frac{1}{\alpha_1} \cdot (e^{\alpha_0 + \alpha_1 T_r} - e^{\alpha_0 + \alpha_1 T_{r-1}}) \tag{3.38}$$

The least-squares method often leads to more appropriate parameters than the Maximum Likelihood method, especially when residual life is estimated.

To estimate next expected failures within confidence bounds it is required to estimate the variance of the model parameters. This is done by calculating the Fischer information matrix for the estimated parameters α_0 and α_1. The Fischer information matrix is given as,

$$\begin{bmatrix} Var(\hat{\alpha_0}) & Cov(\hat{\alpha_0}\hat{\alpha_1}) \\ Cov(\hat{\alpha_1}\hat{\alpha_0}) & Var(\hat{\alpha_1}) \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 I_{\rho_1}}{\partial \hat{\alpha}_0^2} & -\frac{\partial^2 I_{\rho_1}}{\partial \hat{\alpha}_0 \hat{\alpha}_1} \\ -\frac{\partial^2 I_{\rho_1}}{\partial \hat{\alpha}_1 \hat{\alpha}_0} & -\frac{\partial^2 I_{\rho_1}}{\partial \hat{\alpha}_1^2} \end{bmatrix}^{-1} \tag{3.39}$$

where,

$$\begin{aligned} \frac{\partial^2 I}{\partial \hat{\alpha_0}^2} &= -\frac{e^{\hat{\alpha_0}}(e^{\hat{\alpha_1} T_r} - 1)}{\hat{\alpha_1}} \\ \frac{\partial^2 I}{\partial \hat{\alpha_1}^2} &= -\frac{T_r^2 e^{\hat{\alpha_0} + \hat{\alpha_1} T_r} \hat{\alpha}_1^2 - 2T_r e^{\hat{\alpha_0} + \hat{\alpha_1} T_r} \hat{\alpha}_1 + 2e^{\hat{\alpha_0} + \hat{\alpha_1} T_r} - 2e^{\hat{\alpha_0}}}{\hat{\alpha_1^3}} \\ \frac{\partial^2 I}{\partial \hat{\alpha_0} \partial \hat{\alpha_1}} = \frac{\partial^2 I}{\partial \hat{\alpha_1} \partial \hat{\alpha_0}} &= -\frac{T_r e^{\hat{\alpha_0} + \hat{\alpha_1} T_r} \hat{\alpha}_1 - e^{\hat{\alpha_0} + \hat{\alpha_1} T_r} + e^{\hat{\alpha_0}}}{\hat{\alpha}_1^2} \end{aligned} \tag{3.40}$$

The partial derivatives of $\rho_1(t)$ are given by,

$$\begin{aligned} \frac{\partial \rho_1(t)}{\partial \hat{\alpha_0}} &= exp(\hat{\alpha_0} + \hat{\alpha_1} t) \\ \frac{\partial \rho_1(t)}{\partial \hat{\alpha_1}} &= t \cdot exp(\hat{\alpha_0} + \hat{\alpha_1} t) \end{aligned} \tag{3.41}$$

The variance of $\rho_1(t)$ is consequently calculated as,

$$Var(\hat{\rho_1}(t)) = (\frac{\partial \rho_1(t)}{\partial \hat{\alpha_0}})^2 \cdot Var(\hat{\alpha_0}) + (\frac{\partial \rho_1(t)}{\partial \hat{\alpha_1}})^2 \cdot Var(\hat{\alpha_1}) + 2 \cdot (\frac{\partial \rho_1(t)}{\partial \hat{\alpha_0}})(\frac{\partial \rho_1(t)}{\partial \hat{\alpha_1}}) \cdot Cov(\hat{\alpha_0}\hat{\alpha_1}) \tag{3.42}$$

Therefore, $\rho_1(t)$ of equation (3.30) lies between,

$$\hat{\rho_1}(t) - z_\alpha \sqrt{Var(\hat{\rho_1}(t))} \leq \rho_1(t) \leq \hat{\rho_1}(t) + z_\alpha \sqrt{Var(\hat{\rho_1}(t))} \tag{3.43}$$

where $z_\alpha = 1.96$ for a 95% confidence band. It is now possible to calculate any of the descriptive statistics derived in equations (3.31) to (3.33) within statistical bounds with equation (3.43) being known.

Power Law NHPP

The power law NHPP is given by,

$$\rho_2(t) = \lambda \delta (t)^{\delta - 1} \tag{3.44}$$

where $\delta \geq 0$ for repairable systems. By definition the expected number of failures, N, between any two instants, t_2 and t_1, in the process can be obtained by integrating equation (3.44), i.e.

$$E[N(t_1 \rightarrow t_2)] = \lambda(t_2^\delta - t_1^\delta) \tag{3.45}$$

Therefore it is possible to estimate the system's MTBF by,

$$MTBF_{\rho_2}(t_1 \rightarrow t_2) = \frac{(t_2 - t_1)}{\lambda(t_2^\delta - t_1^\delta)} \tag{3.46}$$

In a similar manner, the reliability of the repairable system from t_1 to t_2 can be calculated by,

$$R(t_1 \rightarrow t_2) = e^{-\lambda(t_2^\delta - t_1^\delta)} \tag{3.47}$$

Parameters can be estimated, as before, by either the simple least-squares parameter estimation method or the method of Maximum Likelihood. For the least-squares method, the difference between the observed number of failures and the number of failures expected by $\rho_2(t)$ should be minimized by,

$$min(\hat{\lambda}, \hat{\delta}) : \sum_{i=1}^{r} [E[N(0 \rightarrow T_i)] - N(0 \rightarrow T_i)]^2 \tag{3.48}$$

for the case where T_r is a failure and

$$min(\hat{\lambda}, \hat{\delta}) : \sum_{i=1}^{r-1} [E[0 \rightarrow T_i)] - N(0 \rightarrow T_i)]^2 \tag{3.49}$$

for the case where T_r is a suspension.

The other option is to maximize the log likelihood given by,

$$I_{\rho_2}(\lambda, \delta) = r \cdot ln\lambda + r \cdot ln\delta - \lambda \cdot T_r^\delta + (\delta - 1) \sum_{i=1}^{r} ln \cdot T_i \tag{3.50}$$

such that,

$$max(\hat{\lambda}, \hat{\delta}) : I_{\rho_2}(\lambda, \delta) = I_{\rho_2}(\hat{\lambda}, \hat{\delta}) \tag{3.51}$$

for the case where T_r is a failure. If T_r is a suspension, the likelihood function becomes,

$$I_{\rho_2}(\lambda, \delta) = r \cdot ln\lambda + r \cdot ln\delta - \lambda \cdot T_r^\delta + (\delta - 1) \sum_{i=1}^{r} ln \cdot T_i - \lambda(T_2^\delta - T_1^\delta) \tag{3.52}$$

The least-squares method often leads to more appropriate parameters than the Maximum Likelihood method, especially when residual life is estimated.

The Fisher information matrix for the estimated parameters $\hat{\lambda}$ and $\hat{\delta}$ for determining the power law's confidence bounds is given by,

$$\begin{bmatrix} Var(\lambda) & Cov(\lambda\delta) \\ Cov(\delta\lambda) & Var(\delta) \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 I_{\rho 2}}{\partial\lambda^2} & -\frac{\partial^2 I_{\rho 2}}{\partial\lambda\delta} \\ -\frac{\partial^2 I_{\rho 2}}{\partial\delta\lambda} & -\frac{\partial^2 I_{\rho 2}}{\partial\delta^2} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{r}{\lambda^2} & T_r^\delta \cdot ln(T_r) \\ T_r^\delta \cdot ln(T_r) & \frac{r}{\delta^2} + \lambda T_r^\delta \cdot (lnT_r)^2 \end{bmatrix}^{-1} \tag{3.53}$$

The partial derivatives for $\rho_2(t)$,

$$\frac{\partial\rho_2(t)}{\partial\lambda} = \delta t^{\delta-1} \tag{3.54}$$

$$\frac{\partial\rho_2(t)}{\partial\delta} = \lambda t^{\delta-1} + \lambda\delta t^{\delta-1} ln(t) \tag{3.55}$$

It is now possible to calculate the variance of $\rho_2(t)$ by,

$$Var(\hat{\rho}_2(t)) = (\frac{\partial\rho_2(t)}{\partial\hat{\lambda}})^2 \cdot Var(\hat{\lambda}) + (\frac{\partial\rho_2(t)}{\partial\hat{\delta}})^2 \cdot Var(\hat{\delta}) + 2(\frac{\partial\rho_2(t)}{\partial\hat{\lambda}})(\frac{\partial\rho_2(t)}{\partial\hat{\delta}}) \cdot Cov(\hat{\lambda}\hat{\delta}) \tag{3.56}$$

Therefore, $\rho_2(t)$ of equation (3.44) lies between,

$$\hat{\rho_2}(t) - z_\alpha\sqrt{Var(\hat{\rho_2}(t))} \leq \rho_2(t) \leq \hat{\rho_2}(t) + z_\alpha\sqrt{Var(\hat{\rho_2}(t))} \tag{3.57}$$

where $z_\alpha = 1.96$ for a 95% confidence band.

3.8.2 Residual Life Estimation

The underlying failure process is either described with the log-linear NHPP in §3.8.1 or the power law NHPP in §3.8.1. It is possible to predict the arrival time of the next event. As in the case with non-repairable part, predictions is made based on the conditional expectation of failure. These two cases are discussed separately in the next two sections.

Log-linear NHPP

Only a special case of residual life estimation is considered. Suppose a system has just been put back into operation after its r^{th} failure and it is required to calculate the residual life of the system. The expected arrival time of the $(r+1)^{th}$ failure, given the observation at T_r , is calculated by:

$$E[T_{r+1}|t = T_r] = \frac{ln[(r+1)\hat{\alpha}_1 + exp(\hat{\alpha}_0)] - \hat{\alpha}_0}{\hat{\alpha}_1} \tag{3.58}$$

Since $t = T_r$, it follows that $\mu_{r+1} = T_{r+1} - T_r$.

The upper and lower confidence limits of the residual life estimate can be determined. Due to the complexity of the determining these limits analytically numerical integration is used to simplify the calculations. $\underset{\sim}{T}_{r+1}$ and $\tilde{T}_{r+1}$ are calculated by,

$$\int_{T_r}^{\underset{\sim}{T}_{r+1}} \left[\rho_1(t) + z_\alpha \sqrt{Var(\rho_1(t))}\right] dt = 1 \tag{3.59}$$

and,

$$\int_{T_r}^{\tilde{T}_{r+1}} \left[\rho_1(t) - z_\alpha \sqrt{Var(\rho_1(t))}\right] dt = 1 \tag{3.60}$$

whereafter $\underset{\sim}{\mu}_{r+1} = \underset{\sim}{T}_{r+1} - T_r$ and $\tilde{\mu}_{r+1} = \tilde{T}_{r+1} - T_r$. It is important to note that $\rho_1(t) - z_\alpha \sqrt{Var(\rho_1(t))}$ is only defined for values greater than zero. In cases where $\rho_1(t)$ does not fit the data well, the integral in equation (3.60) will not converge to 1. In these cases it is not possible to quantify the upper limit.

Power Law NHPP

Similar to the log-linear NHPP scenario a system has just been put back into operation after its r^{th} failure and it is required to calculate the residual life of the system. The expected arrival time of the $(r+1)^{th}$ failure, given the observation at T_r , is calculated by:

$$E[T_{r+1}|t = T_r] = \left(\frac{1 + \lambda T_r^\delta}{\lambda}\right)^{1/\delta} \tag{3.61}$$

Since $t = T_r$, it follows that $\mu_{r+1} = T_{r+1} - T_r$.

The upper and lower confidence limits of the residual life estimate can now be determined. $\underset{\sim}{T}_{r+1}$ and $\tilde{T}_{r+1}$ are calculated by,

$$\int_{T_r}^{\underset{\sim}{T}_{r+1}} \left[\rho_2(t) + z_\alpha \sqrt{Var(\rho_2(t))}\right] dt = 1 \tag{3.62}$$

and,

$$\int_{T_r}^{\tilde{T}_{r+1}} \left[\rho_2(t) - z_\alpha \sqrt{Var(\rho_2(t))}\right] dt = 1 \tag{3.63}$$

whereafter $\underset{\sim}{\mu}_{r+1} == \underset{\sim}{T}_{r+1} - T_r$ and $\tilde{\mu}_{r+1} = \tilde{T}_{r+1} - T_r$. Similar to the log-linear format $\rho_2(t) - z_\alpha \sqrt{Var(\rho_2(t))}$ is only defined for values greater than zero, with cases where $\rho_1(t)$ does not fit the data well, the integral in equation (3.63) will not converge to 1 and it will not be possible to quantify the upper limit.

3.8.3 Long Term Cost Optimisation

The methodology to optimise the long term maintenance cost of repairable systems is fundamentally different from non-repairable parts. It is assumed that during a repairable system's existence, minimal repairs are performed after each event at an average cost C_m to restore the system to the same state as just before the event. At the end of the repairable system's existence (typically when minimal repairs need to be performed too frequently), the current system is discarded and replaced by a new, similar system at a cost C_s. To optimise long term cost, the instant of complete system replacement needs to be calculated. This instant is either expressed by N*, which is the optimal number of minimal repairs or, T*, which is the optimal global time.

Log-linear NHPP

The cost per unit time in terms of t is given by,

$$\begin{aligned} C_R(t) &= \frac{(E[N(0 \to t)] - 1) \cdot C_m + C_s}{t} \\ &= \frac{(1/\alpha_1 \cdot [exp(\alpha_0 + \alpha_1(t) - exp(\alpha_0)] - 1) \cdot C_m + C_s}{t} \end{aligned} \tag{3.64}$$

and T* is found where,

$$min(t) : C_R(t) = C_R(T*) \tag{3.65}$$

Similarly, the cost in terms of N is given by,

$$C_R(N) = \frac{\alpha_1((N-1) \cdot C_m + C_s)}{ln(\alpha_1 N - exp(\alpha_0)) - \alpha_0} \tag{3.66}$$

and N* is found where,

$$min(N) : C_R(N) = C_R(N*) \tag{3.67}$$

In both equations (3.66) and (3.67), N should be integer numbers.

Power Law NHPP

The cost per unit time in terms of t is given by,

$$\begin{aligned} C_R(t) &= \frac{(E[N(0 \to t)] - 1) \cdot C_m + C_s}{t} \\ &= \frac{(\lambda t^{\delta} - 1) \cdot C_m + C_s}{t} \end{aligned} \tag{3.68}$$

and T* is found where,

$$min(t) : C_R(t) = C_R(T*) \tag{3.69}$$

Similarly, the cost in terms of N is given by,

$$C_R(N) = \frac{(N-1) \cdot C_m + C_s}{(N/\lambda)^{1/\delta}} \tag{3.70}$$

and N* is found where,

$$min(N) : C_R(N) = C_R(N*) \tag{3.71}$$

In both equations (3.70) and (3.71), N should be integer numbers.

***Example* 3.8.** *Below are interarrival times to events recorded from a circulating pump. In this case there are no failure-type indicators because the pump was never replaced but simply restored to a BAO condition and put back into service.*

304, 120, 119, 131, 60, 26, 23, 36, 30, 29, 18, 15, 20, 11

First it is required to test for a trend in the data with Laplace's test described in equation (3.72), i.e.

$$\begin{aligned} U &= \frac{\frac{9506}{13} - \frac{942}{2}}{942\sqrt{\frac{1}{12(14-1)}}} \\ &= 3.4504 \end{aligned}$$

This result shows that there is a strong indication of reliability degradation of the pump. Repairable systems theory is therefore appropriate. For illustrative purposes, the data set is modeled with both the log-linear NHPP and the power law NHPP in the sections below.

***Example* 3.9.** *To fit the log-linear NHPP, $\rho_1(t)$ of equation (3.30) to the data set, the least-squares method of equation (3.34) is used. This yields,*

$$\rho_1(t) = exp(-6.809 + 0.004214t) \tag{3.72}$$

In Figure 3.13 the observed number of failures are compared to the number of failures expected by $\rho_1(t)$ as given by equation (3.31). It is clear that (3.72) is a good model of the actual situation.
If we would like to calculate the MTBF of the pump over the observed period, equation (3.32) is applicable:

$$\begin{aligned} MTBF_{\rho_1}(0 \to 942) &= \frac{0.004212 \cdot (942 - 0)}{exp(-6.808 + 0.004212 \cdot 942) - exp(-6.808 + 0.004212 \cdot 0)} \\ &= 69.25 \text{ days} \end{aligned}$$

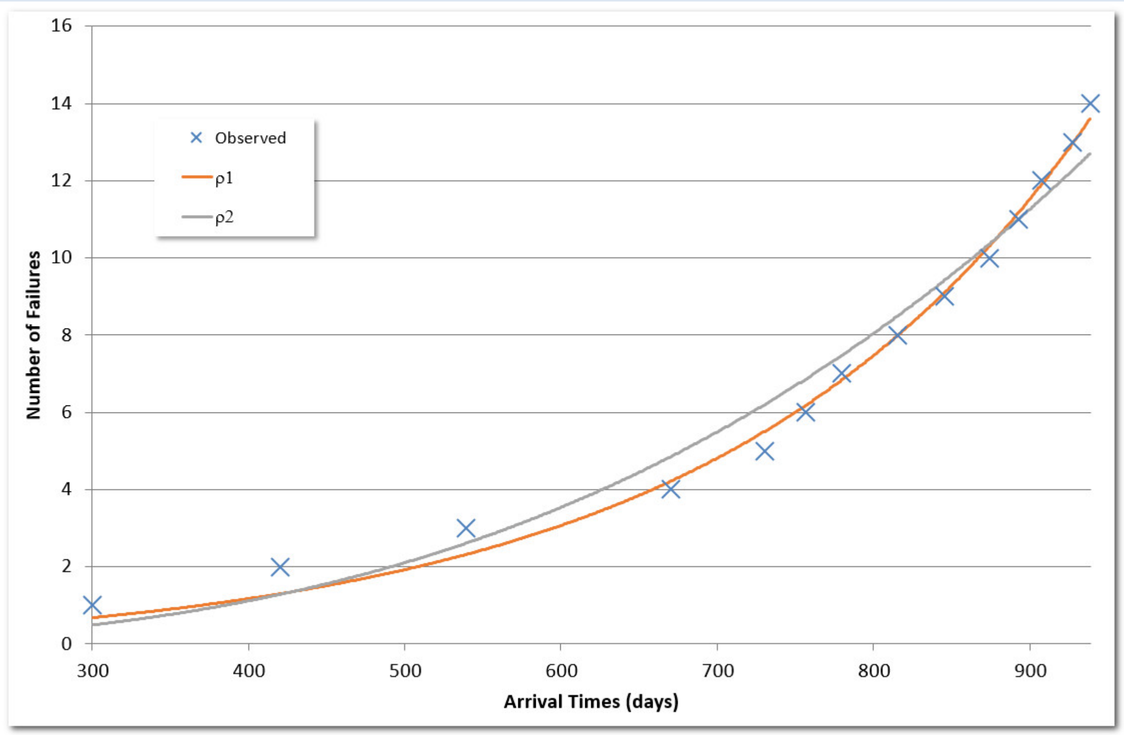


Figure 3.13: Comparison of $E[N(0 \to t)]$ by $\rho_1(t)$ and $\rho_2(t)$ vs observed $N(0 \to t)$

12

Suppose we have just put the pump back into operation after observing failure r = 14. The expected arrival time of the next failure is given by equation (3.58), i.e.

$$\begin{aligned} E[T_{15}|T_{14}=942] &= \frac{ln[(14+1)\cdot 0.004212 + exp(-6.808)] + 6.808}{0.004212} \\ &= 964 \textit{ days} \end{aligned}$$

and it follows that $\mu_{15} = 964.76 - 942 = 22.83$ *days.*
The confidence limits for this estimate can now be determined. The Fischer information matrix of equation (3.39) is,

$$\begin{bmatrix} Var(\alpha_0) & Cov(\alpha_0\hat{\alpha_1}) \\ Cov(\alpha_1\alpha_0) & Var(\alpha_1) \end{bmatrix} = \begin{bmatrix} 1.060266 & -0.0013652 \\ -0.001365 & 0.000001889 \end{bmatrix} \tag{3.73}$$

Equations (3.59) and (3.60) are used to calculate confidence bounds around the estimate. Equation (3.60) does not converge to 1 and therefore only a lower confidence limit is specified,

$$E[T_{15}|T_{14}=942] = 951.24 \textit{ days} \tag{3.74}$$

which implies that $\mu_{15} = 9.24$ *days. Figure 3.14 illustrates the divergence of the upper*

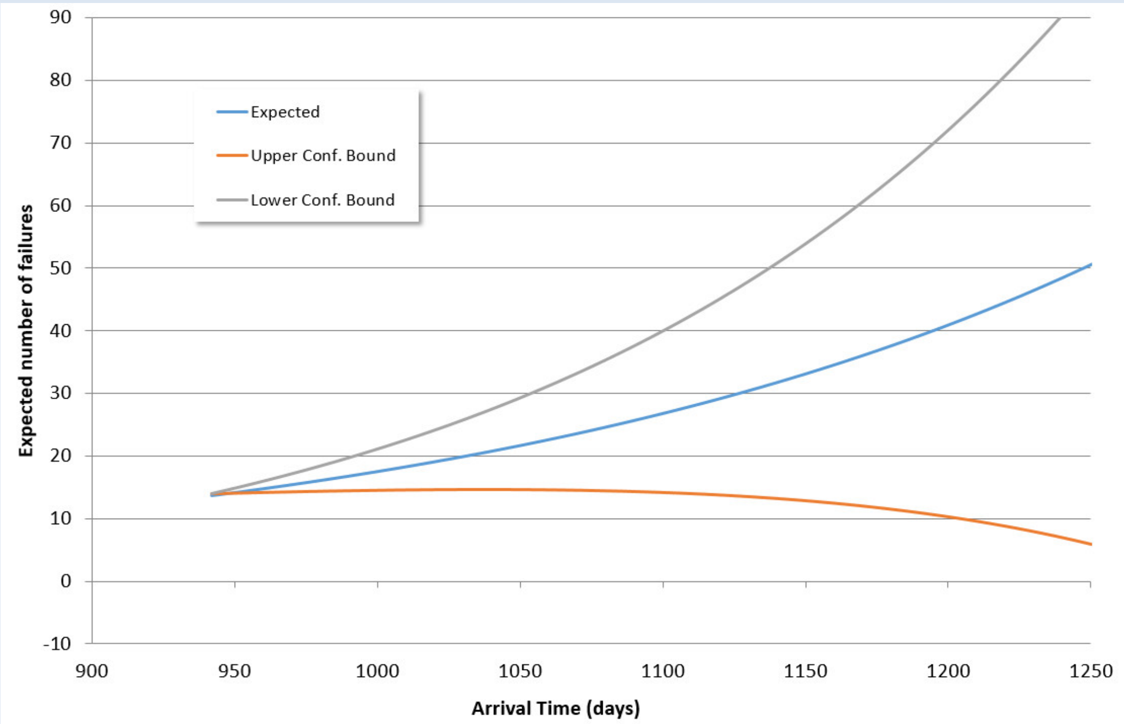


Figure 3.14: Number of failures expected by $\rho_1(t)$ within confidence bounds

***Example* 3.10.** *Assume it costs R 30 000 to replace the entire pump system and R 2 000 on average to perform minimal repair to the pump. Equations (3.64) and (3.66) show that the optimal system replacement instant in terms of time is* $T* = 755$ *days where* $C_R = R53.09$ *per day. In terms*

14

of the number of minimal repairs, the optimal system replacement instant is at $N = 7$ times where $C_R = R54.49$ per day. This system has clearly been operated beyond its most economical replacement time.*

***Example* 3.11.** *To fit the power law NHPP, $\rho_2(t)$ of equation (3.44) to the data set, the least-squares method of equation (3.48) is used. This yields,*

$$\rho_2(t) = (3.68e - 10) \cdot 2.8709 \cdot t^{1.8709} \tag{3.75}$$

In Figure 3.13 the observed number of failures are compared to the number of failures expected by $\rho_2(t)$ as given by equation (3.75). Equation (3.75) is also a good representation of the actual situation compared to $\rho_1(t)$.

If we would like to calculate the MTBF of the pump over the observed period, equation (3.46) is applicable, i.e.

$$\begin{aligned} MTBF_{\rho_2}(0 \to 942) &= \frac{(942 - 0)}{(3.68e - 8) \cdot (942^{2.8709} - 0^{1.8709}} \\ &= 70.58 \text{ days} \end{aligned}$$

Suppose we have just put the pump back into operation after observing failure r = 14. The expected arrival time of the next failure is given by equation (3.61), i.e.

$$\begin{aligned} E[T_{15}|T_{14} = 942] &= \left(\frac{1 + (3.68e - 8)942^{2.8709}}{(3.68e - 8)}\right)^{1/1.8709} \\ &= 967.21 \text{ days} \end{aligned}$$

and it follows that $\mu_{15} = 967.21 - 942 = 25.21$ days.
The confidence limits around this estimate is calculated. The Fischer information matrix of equation (3.53) is,

$$\begin{bmatrix} Var(\alpha_0) & Cov(\alpha_0\hat{\alpha}_1) \\ Cov(\alpha_1\alpha_0) & Var(\alpha_1) \end{bmatrix} = \begin{bmatrix} 3.76262e - 19 & -1.08476e - 10 \\ -1.08476e - 10 & 0.03286877 \end{bmatrix} \tag{3.76}$$

Equations (3.62) and (3.63) are used to calculate confidence bounds around the estimate,

$$\begin{aligned} E[T_{15}|T_{14} = 942] &= 954.64 \text{ days} \\ E[\tilde{T}_{15}|T_{14} = 942] &= 986.16 \text{ days} \end{aligned}$$

which implies that $\tilde{\mu}_{15} = 12.64$ days and $\tilde{\mu}_{15} = 44.16$ days. The number of failures expected by $\rho_2(t)$ within confidence bounds are shown in Figure 3.15 illustrating the reason why equation (3.63) converges to 1 for $\rho_2(t)$.

16

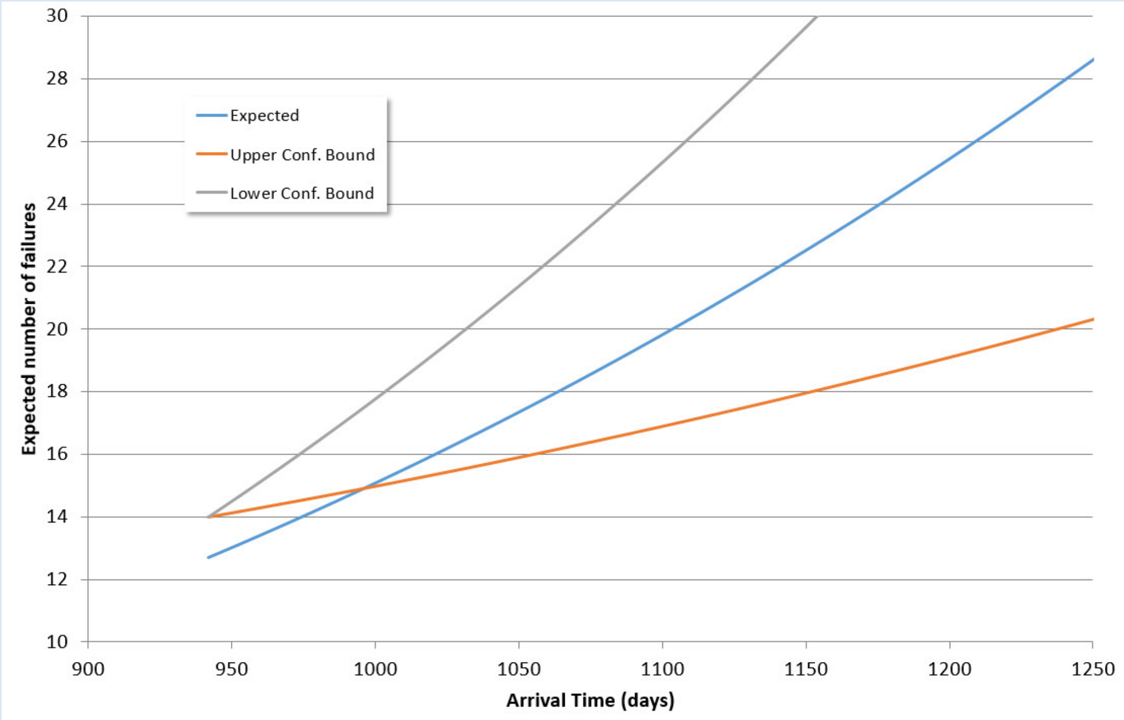


Figure 3.15: Number of failures expected by $\rho_2(t)$ within confidence bounds

***Example* 3.12.** *Assume it costs R 30 000 to replace the entire pump system and R 2 000 on average to perform minimal repair to the pump. Equations (3.68) and (3.70) show that the optimal system replacement instant in terms of time is $T* = 784$ days where C_R = R 54.83 per day. In terms of the number of minimal repairs, the optimal system replacement instant is at $N* = 7$ times where C_R = R 54.86 per day. This system has clearly been operated beyond its most economical replacement time.*

3.9 Availability

For repairable systems the "classic" definition of reliability only applies to the time of the first failure. Therefore the reliability equivalent for a repairable system is *availability*. Availability is defined as the probability that an item will be available when required, or as the proportion of total time that the item is available for use. The availability is a function of the ROCOF, (refer to §3.2.2) and the repair rate. The proportion of total time that the item is available (functional) is the *steady-state availability*. For a simple unit, with a constant ROCOF, $\lambda = \frac{1}{\text{MTBF}}$ and a constant mean repair rate, μ, where $\mu = \frac{1}{MTTR}$, the steady state availability is equal to:

$$A = \frac{MTBF}{MTBF + MTTR} = \frac{\mu}{\lambda + \mu} \tag{3.77}$$

Another availability variant is the instantaneous availability, $A(t)$ or probability that the item will be available at time, t. For a simple unit this is equal to:

$$A(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}, \tag{3.78}$$

and which approaches the steady-state availability as t becomes large.

Availability is also affected by redundancy. If standby systems can be repaired while the primary system provides the required function, the overall system availability can be improve. Table 3.7 shows the steady-state availability functions for some simple system configurations. The gains in reliability and availability as a result of redundancy is clear. These functions are for relatively simple situations where the failure rate is constant.

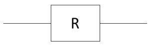
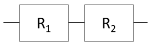
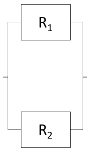
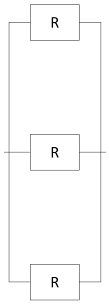
Reliability Configuration		**Steady-state availability, A, repair rate, μ, ROCOF, λ**	**General steady-state availability, A, n blocks**
	–	$\frac{\mu}{\lambda+\mu}$	–
	–	$\frac{\mu_1\mu_2}{\mu_1\mu_2+\mu_1\lambda_2+\mu_2\lambda_1+\lambda_1\lambda_2}$	$\prod_{i=1}^{w}\frac{\mu_i}{\lambda_i+\mu_i}$
	Active	$\frac{\mu^2+2\mu\lambda}{\mu^2+2\mu\lambda+2\lambda^2}$[1]	$1-\prod_{i=1}^{w}\frac{\lambda_i}{\lambda_i+\mu_i}$
	Standby	$\frac{\mu^2+\mu\lambda}{\mu^2+\mu\lambda+\lambda^2}$	–
	Active 1/3	$\frac{\mu^3+3\mu^2\lambda+6\mu\lambda^2}{\mu^3+3\mu^2\lambda+6\mu\lambda^2+6\lambda^3}$	As above (active)
	Active 2/3	$\frac{\mu^3+3\mu^2\lambda}{\mu^3+3\mu^2\lambda+6\mu\lambda^2+6\lambda^3}$	$1-\frac{1}{(\lambda+\mu)^w}\sum_{i=0}^{k-1}\binom{w}{i}\mu^i\lambda^{w-i}$
	Standby 1/3	$\frac{\mu^3+\mu^2\lambda+\mu\lambda^2}{\mu^3+\mu^2\lambda+\mu\lambda^2+\lambda^3}$	–

Table 3.7: Steady-state availability functions for some system configurations

It is evident from table 3.7 even with constant failure rates the steady-state availability quickly becomes statistically complex for $n > 3$. Refer to O'Connor and Kleyner (2012, 148) and Myers et al. (1964) for details about availability for more complex systems.

***Example* 3.13.** *Recall the missile system in example 2.7. Determine the steady-state availability*

[1] $\lambda_1 = \lambda_2 = \lambda$. Assumes series repair, that is single repair team.

of the system, excluding the missiles, if the mean repair time for all units is 2 hours.

$$
\begin{aligned}
\mu &= \frac{1}{MTTR} = \frac{1}{2} = 0.5 \\
\lambda_{LG} &= \frac{1}{750} = 0.0013 \text{ (from example 2.7)} \\
\lambda_R &= 0.001 \text{ (from example 2.7)} \\
A_R &= \frac{\mu^2 + \mu\lambda}{\mu^2 + \mu\lambda + \lambda^2} \text{ (from table 3.7)} \\
&= \frac{0.5^2 + 0.5 \cdot 0.001}{0.5^2 + 0.5 \cdot 0.001 + 0.001^2} \\
&= 0.999996 \\
A_{LG} &= \frac{\mu}{\lambda + \mu} \\
&= \frac{0.5}{0.0013 + 0.5} \\
&= 0.9974 \\
A_S &= A_R A_{LG} \\
&= 0.9974
\end{aligned}
$$

This example is an illustration of how availability calculations can be used for performing sensitivity analysis to different system designs. For example, a 20% reduction in the MTBF of the launch guidance systm would have far greater impact on the system reliability than a similar reduction in the MTBF of the two radars.

18

Acronyms

C

CAD Computer-aided Design
CAE Computer-aided Engineering
CFD Cumputational Fluid Dynamics
CFR Constant Failure Rate
CMMS* Computerised Maintenance Management System
COTS Commercial Off The Shelf

D

DfR Design for Reliability
DOE Design of Experiments

E

EAMS* Enterprise Asset Management System
EDA Electronic Design Automation

F

FAA Finite Element Analysis
FMEA Failure Mode and Effects Analysis
FOM Force of Mortality
FRACAS Failure Reporting, Analysis and Corrective Action System
FTA Fault Tree Analyis

H

HALT Highly Accelerated Life Testing
HASS Highly Accelerated Stress Screening
HPP Homogeneous Poisson Process

I

IID Independent and Identically Distributed
ISO International Organisation for Standardisation

L

LCC Life Cycle Cost

M

MTBF Mean Time Between Failure
MTTF Mean Time to Failure

N

NHPP Non-homogeneous Poisson Process

Q

QA Quality Assurance
QFD Quality Function Deployment

R

RBD Reliability Block Diagrams
RCM Reliability Centered Maintenance
ROCOF Rate of occurance of failures

S

SPC Statistical Process Control

T

TQM Total Quality Management

References

T.W. Anderson and D.A. Darling. A test of goodness of fit. *ournal of the American Statistical Association*, pages 765–769, 1954.

H. Ascher and H. Feingold. *Repairable Systems Reliability*. Marcel Dekker, Inc., 1984.

H.A. Ascher and C.K. Hansen. Spurious Exponentiality Observed When Incorrectly Fitting a Distribution to Nonstationary Data. *IEEE Transactions on Reliability*, 47(4):451–459, 1998.

Z. W. Birnbaum. On the importance of different components in a multicomponent system. Technical report, Laboratory of Statistical Research, University of Washington, 1968.

J.L. Coetzee. *Maintenance*. Trafford Publishing, 2004.

J. Cohen. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum, 1988.

D.R. Cox and P.A. Lewis. *The statistical analysis of series of events*. Methuen, 1966.

E.A. Elsayed. *Reliability Engineering*. Johan Wiley & Sons, New Jersey, 2012.

R. L. Finney and G. B. Thomas. *Calculus*. Addison Wesley, Warrendale, 2 edition, 1993.

A. Gretton, K. Fukumizu, C. Teo, L. Song, B. Scholkopf, and A. Smola. A kernel statistical test of independence. In *Advances in neural information processing systems*. 2007.

WW Hines, DC Montgomery, DM Goldsman, and CM Borror. *Probability and Statistics in Engineering*. John Wiley & Sons, New Jersey, fourth edition edition, 2003.

A.K.S. Jardine and A.H.C. Tsang. *Maintenance, Replacement, and Reliability*. Taylor & Francis, 2006.

Renyan Jiang. *Introduction to Quality and Reliability Engineering*. Springer, 2015.

M Kaminskiy and V Krivtsov. G-Renewal process as a model for statistical warranty claim prediction. In *Proceedings of the annual reliability and maintainability symposium (RAMS)*, pages 1668–1674, Shenzhen, China, 2000.

E. Lee and J. Wang. *Statistical Methods for Survival Data Analysis*. Wiley, Oxford, 2013.

P.A. Lewis and D.W. Robinson. Testing for a monotone trend in a modulated renewal process. *Reliability and Biometry*, pages 163–182, 1974.

H.B. Mann. Nonparametric tests against trend. *Econometrica*, pages 245–259, 1945.

JI McCool. *Using the Weibull Distribution*. John Wiley & Sons, New Jersey, 2012.

DNP Murthy, M Rausand, and T Østerås. *Product reliability: specificaiton and performance*. Springer, 2009.

R. H. Myers, K. L. Wong, and H. M. Gordy. *Reliability Engineering for Electronic Systems*. John Wiley & Sons, New York, 1964.

P.D.T. O'Connor and A. Kleyner. *Practical Reliability Engineering*. John Wiley & Sons Ltd, Chichester, UK, fifth edition edition, 2012.

M. Rausand and A. Hoyland. *System reliability theory: Models, statistical methods, and applications*. John Wiley & Sons, New Jersey, 2004.

Reliawiki.org. Reliability Growth and Repairable System Analysis, 2016. URL http://www.reliawiki.org/index.php/RGA_Overview.

D. J. Sherwin and A. Bossche. *The Reliability, Availability and Productiveness of Systems*. Springer, New York, second edition, 1993.

P.J. Vlok. Introduction to Elementary Statistical Analysis of Failure Time Data: Long Term Cost Optimisation and Residual Life Estimation. Technical report, Stellenbosch University, 2014.

Ronald E. Walpole, Raymond H. Myers, and Sharon L. Myers. *Probability and statistics for engineers and scientists*. Prentice Hall, New Jersey, 6th edision edition, 1998.

G. Wasserman. *Reliability verification testing, and analysis in engineering*. Marcel Dekker Inc., New York, 2003.