Faculdade de Engenharia da Universidade do Porto

The advanced reliability and maintainability analysis using simulation tools

Daniel Augusto Estácio Marques Mendes Gaspar

Programa Doutoral em Engenharia Mecânica

Supervisor: Professor Luis Andrade Ferreira

Junho, 2019

The advanced reliability and maintainability analysis using simulation tools

Daniel Augusto Estácio Marques Mendes Gaspar

Programa Doutoral em Engenharia Mecânica

Junho, 2019

Resumo

As novas ferramentas de *Inteligência Artificial (AI), Machine Learning (ML) e Predictive Analysis (PA)* estão a mudar a forma como olhamos para uma das áreas clássicas da gestão das operações: a manutenção.

A aplicação de modernas técnicas de simulação e algoritmos, computacionalmente poderosas, permitem trabalhar os dados de avarias (histórico) de forma a estimar os parametros da fiabilidade e deste modo optimizar e melhorar a manutenção de equipamentos/componentes.

Este trabalho de doutoramento, tem como objectivo a aplicação da simulação *Monte Carlo* na fiabilidade de equipamentos e o desenvolvimento de novas metodologias e ferramentas de simulação, com o uso de algoritmos desenvolvidos em duas linguagens de programação: *R e Python.*

É apresentada uma metodologia para avaliar e testar os algoritmos de simulaçao de fiabilidade, nomedamente de dados censurados. Foram desenvolvidos algoritmos de simulação de dados censurados à direita, tipo I (fixo e aleatéorio) e dados censurados à direita tipo II. A metodologia inclui um conjunto de testes de hipóteses que permitem avaliar se os dados simulados são gerados de uma forma correta, robusta e não-enviesada .

Para avaliar e estimar os parâmetros dos dados históricos da manutenção (escassos e censurados), desenvolveu-se analiticamente o método de máxima verosimilhança para a distribuição de Weibull e toda a derivação do método EM- Expectation-Maximization para a estimação dos parâmetros de distribuição Weibull com dados censurados.

Desenvolveu-se uma metodologia para para simular e calcular o indicador "importância dos componentes" , bem como a aplicação da simulação e do conceito de sistemas lineares consecutivos “k-out-of-n”.

Desta forma, pode-se melhorar a gestão da manutenção, principalmente com um planeamento dos intervalos de manutenção mais rigoroso e com o risco aceitável.

Foi desenvolvido um caso de estudo numa empresa química e com dados históricos. Os resultados deste trabalho permitiram, por um lado, validar as metodologias e ferramentas desenvolvidas, e por outro, comparar cenários e concluir que existe margem para o aumento dos intervalos de manutenção em equipamentos/componentes sem dimuir a fiabilidade e/ou aumentar o grau do risco.

Estas ferramentas, bem como as metodologias desenvolvidas, poderão ser aplicadas noutras indústrias, mas também, podem ser usadas dentro dos algoritmos ou em simultâneo com as novas ferramentas *Inteligência Artificial (AI), Machine Learning (ML) e Predictive Analysis (PA).*

Abstract

The new Artificial Intelligence *(AI)*, Machine Learning *(ML)* and Predictive Analysis *(PA)* tools are changing the way we look at one of the classic operations management areas: maintenance.

Applying modern simulation techniques and algorithms, which are computationally powerful, allows working the fault data (historical data) in order to estimate the parameters of reliability and optimize and improve the maintenance of the equipment / components.

The objective of this PhD work is to apply the Monte Carlo simulation for equipment reliability and develop new methodologies and simulation tools, using algorithms developed in two programming languages: R and Python.

A methodology is presented to evaluate and test the reliability simulation algorithms, namely censored data. Right type I (fixed and random) and right type II censored data simulation algorithms were developed. The methodology includes a set of hypothesis tests that allow evaluating if the simulated data are generated in a correct, solid and non-skewed way.

In order to evaluate and estimate the parameters of the historical maintenance data (scarce and censored), the maximum likelihood method for Weibull distribution and the entire derivation of the EM-Expectation-Maximization method for the estimation of the Weibull distribution parameters with censored data were developed analytically.

A methodology was developed to simulate and calculate the "component importance" indicator, as well as for application of the simulation and of the concept of consecutive "k-out-of-n" linear systems. With these tools, maintenance management can be improved, especially through stricter maintenance interval planning and acceptable risk.

A case study was developed in a chemical company with historical data. The results of this work allowed, on the one hand, validating the methodologies and tools developed and, on the other hand, comparing scenarios and concluding that there is space to increase maintenance intervals of equipment/components without reducing reliability and/or increasing the degree of risk.

These tools, as well as the methodologies developed, can be applied in other industries, but can also be used within the algorithms or simultaneously with Artificial Intelligence *(AI)*, Machine Learning *(ML)* and Predictive Analysis *(PA)* tools.

Acknowledgements

I gratefully thank to my supervisor, Professor Luis A. Ferreira for giving me the opportunity to do my PhD, and for their excellent advice, insight, support and encouragement during the course of my doctoral studies.

My gratitude to Professor José Torres Farinha and Professor José Sobral who served as the members of my research committee, and provided me with valuable comments on and insights into both the research and the presentation of the thesis.

I also benefited from the advice, experience, and friendship of a great group of friends and colleagues at IPV during these past years. My thanks especially go to my colleague José Luis Silva , for his exceptionally assistance and comments, which helped me to improve my research papers and presentation of the thesis and my colleague Márcio Ferreira for help me in mathematics subjects.

Finally, I would like to express my deepest gratitude to my family, especially my Parents, who supported me and always believed in me. They have been my greatest mentors. His enthusiasm and encouragement gave me strength to succeed.

Daniel Gaspar

"Scientific predictability also raises the question of the scientist's ethical responsibilities. His conclusions must be guided by respect for truth and an honest acknowledgment of both the accuracy and the inevitable limitations of the scientific method. Certainly this means avoiding needlessly alarming predictions when these are not supported by sufficient data or exceed science's actual ability to predict. "

Joseph Ratzinger, Benedict XVI

Contents

List of Figures

List of Tables

List of program codes

Abbreviations and Symbols

Abbreviations

ABM	Agent Based Modeling
CDF	Cumulative Distribution
CI	Criticality Importance
DES	Discrete Event System
EM	Expectation - Maximization
FOMs	Figure of Merits
FT	Fault Tree
FV	Fussell-Vesely's measure
K-R	Kamat-Riley method
KLMC	Kim-Lee Monte Carlo method
LCLs	Limit Confidence Lower bounds
M&S	Modelation and Simulation
MCm	Monte Carlo methods
MLE	Maximum Likelihood Estimation
MTBF	Mean Time Between Failure
OREDA	Offshore Reliability Data Handbook
PDF	Probability Density function
PRSD	Percentage Relative Standard Deviation
R-M	Rice-Moore method
RAW	Risk Achievement Worth
RBD	Reliability Block diagrams
RNG	Random Number Generator
RRW	Risk Reduction Worth
WWW	*World Wide Web*

Symbols

β	Shape factor from weibull, Gamma
α	Scale factor from weibull, Gamma
σ	standard deviation
$\lambda(t)$	failure rate
μ	Maximum Likelihood Estimation
$\%C$	Percentage of data censored
$R(t)$	Reliability function
$S(t)$	Survival function
$h(t)$	Hazard Function
$U(0,1)$	Uniform random variable
H_0	Null hypothesis
H_1	Alternative hypothesis
α	target test level or level of significance
p	p-value
D_α	critical value Kolmogorov-smirnov
Z_0	Test statistics to Runs test/Wald-Wolfowitz
S	Test statistics Mann-Kendall
n	number of sample
M	number of cycle simulation
t_i	Random times
t_c	time censoring
T_{cens}	Vector censor data
ϵ	error parameter
$L(\theta; x, \delta)$	likelihood function
δ_i	Censorship vector
X_i	$min(T_i, C_i)$
$Q(\theta \mid \theta^{(K)})$	Step E - expectation
θ	the parameters that characterize the distribution.
$\theta^{(0)}$	Initial solution
x_i	the state of component
$\phi(x)$	the *structure function* of the system
P_j	minimal path set
ρ_j	minimal path series structure
K_j	minimal cut set
κ_j	minimal parallel cut structure
$(1_i, x)$	*critical* path vector for component i
$(\cdot_i, x)$	possible state vectors
z_i	minimal path vector of ϕ.
y_i	veto vector of ψ
ψ_{yi}	system of veto vectors (Park)
i	index for components; i = 1,2, ..., n

Chapter 1

Introduction

Reliability and maintainability analysis using simulation tools is the theme of this dissertation. The area and application of simulation is very general and can be based on computers, scale models, prototypes, etc. The maintenance field has to develop in order to increase its scope and diversify intervention policies. This evolution must be accompanied by the recognition of its importance in all business activities and by interpenetration with other functional areas.

There are many reasons guiding the search for knowledge and the need to perceive what surrounds us. Solving problems is part of human nature. The first method of learning about nature was experimentation, and for thousands of years it was the only one available, and remains to this day one of the main ways of solving problems. In experimentation, there are two entities under evaluation: the object studied and the framework of the conditions that defines how the experiment takes place.

Sometimes, the problems to be solved are so complex that experimentation may prove to be an inadequate methodology due to ethical, risk or cost issues. In other cases, the solution is simply impossible to put into practice. Science has been able to find ways to address these issues, one of which is to abstract ourselves from the problem and think of it using a model of a problem. Although different modelling techniques have been proposed over the last few centuries, the formalism of Newton and Leibniz's differential equations has been one of the tools of choice for modelling and problem solving. These equations provide a mathematical (analytical) way of studying the entity in question.

The techniques used in this type of problem solving are analytical in the sense that they are symbolic and based on reasoning, providing generic solutions for the problems faced. An abstraction of the entity is made and the resulting model is then used to evaluate the system. This abstraction is made with loss of information (simplifications), but it represents the behaviour of the entities, allows analysis and is sufficient to demonstrate the properties of the suggested model. It is essential for the results of the model to be in agreement with those observed in the entity, in which case they are consider as valid.

In the case of models with a greater degree of complexity, it is necessary to use differential equations and other numerical methods, approaching the equation through the discretization of its behaviour (calculated at various points in time). Results will be less accurate than analytical models, but will provide values with sufficient accuracy to study the problem under analysis.

The process begins by obtaining experimental data from the entity under study, then mathematically modelling the observed behaviour. However, due to the complexity involved, the equations have to be solved by numerical approximation. With the massification of the use of computers, this is achieved recursively, calculating the values of the state variables at specific moments and using these values as the basis for the next iteration.

The model should provide a solution to the problem; however, this is an approximation and implies a loss in precision when compared to analytical models.

It was noted that it is necessary to introduce a step to verify the accuracy of the results obtained by numerical approximation. The results have to be compared with experimental data, so that the model emulates reality as best as possible. For centuries, these approaches have allowed for very important advances in the science and technology area.

The evolution of our knowledge and the invention of advanced devices (e.g. control systems, intelligent manufacturing, traffic management and control) made it impossible to continue to exclusively use analytical methods for problem solving. These are analytically impossible and numerically dicult to evaluate without oversimplifying the models and can result in solutions that are incoherent with reality.

Starting in the 1940's, computers provided alternative ways of dealing with these issues, since they are well equipped to apply approximation techniques, eliminate human errors, and solve problems at a higher speed. Thus, since the advent of computing, the numerical models have been converted into computational solutions (also called computational simulations). A computational simulation cycle can also be characterized, the difference being that the model is executed by specialized devices. Verification continues to be necessary, as precision limitations can cause erroneous results and diverge from expected solutions.

Simulation enables experiments with virtual environments, advancing the level of analysis of natural and artificial applications to unprecedented levels in scientific history and allowing the design and analysis of complex applications. Simulation also provides cost-effective and risk-free training solutions when compared to experimentation. In sum, simulation is a mature and economical problem solving approach, which will continue to develop with the increase in available computational power.

1.1 Motivation

Simulation and its methods, such as the Monte Carlo method, are widely used in areas such as operations management, logistics, engineering and economics. In this thesis, the goal is to contribute to the development of the scientific field of maintenance in Portugal

and also, if possible, to contribute, overall, with new ideas and solutions to the reliability area.

Simulation and its methods, such as the Monte Carlo method, is widely used in various scientific fields and in areas such as operations management, logistics and economics. Applying simulation methods for reliability was developed mainly by mathematicians with deductions and highly sophisticated demonstrations applied to very specific cases. These models usually used advanced statistical and mathematical tools with excellent theoretical results. However, it is precisely this intensive use of advanced and difficult to interpret tools that require a long time to learn all the steps of the deductions, making their practical application difficult and creating a division from the real and industrial world. This contributes to the lack of interest of the engineering researchers and, particularly, of the maintenance/reliability technicians of the companies.

The study on the reliability of equipment and infrastructures has faced complex systems, that is, they are not reduced to a simple series-parallel system, usually with little data on failures making statistical analysis difficult. On the other hand, much of the reliability literature is centred on component testing where conditions are easily controllable and the data obtained is reliable and organized.

Industrial managers and scientific reliability researchers find it very difficult to be useful to each other. The maintenance sector needs expeditious conclusions and analysis and likes practical tools that are not very complex.

A maintenance manager needs practical and direct tools and academics and researchers need reliable and organized historical data and time for their analysis and conclusions. It is difficult for a manager (other than engineering) to value reliable research or optimization work within the maintenance area unless the cost reduction is significant.

In Portugal, there seems to be little use of simulation for reliability and the use of simulation of scenarios for the planning and optimization of maintenance programs also seems to be scarce.

This thesis intends to increase the use of advanced reliability techniques in Portugal, namely using a methodology that proposes, in the end, using the various methods and simulation software available in the market.

1.2 Objectives

As this work was carried out within the scope of a doctoral dissertation in mechanical engineering (industrial maintenance), there was the constant care to fit the objectives and methodologies used in this area of knowledge.

Being aware that the time available would be limited, efforts were made to delimit the objectives to ensure a high degree of rigour and formalism in order to achieve effective results. There were two purposes that guided the work and that were confused in many situations: research and learning. First, the objectives that increase our knowledge and,

second, the objectives that would carry out unprecedented research work, clearly valuable for the maintenance management, information systems and simulation areas.

The maintenance area encompasses many fields of Science, from equipment management to specific engineering fields, such as mechanic of fracture, corrosion, electronic devices, among others. One of the main areas of maintenance is reliability in the operation of the equipment and in development and testing.

The use of statistical models and computational algorithms has allowed for great development of reliability over the last decades; however, there are still many fields of difficult resolution, mainly the connection between the models developed academically and the reality of the operation and maintenance of equipment within the companies.

More concretely, the objectives of the work can be detailed through the following topics:

- Develop knowledge of reliability with the use of modern engineering techniques applied to the maintenance of productive equipment;
- Develop and contribute to the use of simulation techniques in maintenance engineering;
- Understand the relationship between organizational concepts, maintenance management models and information systems and data collection models for reliability and simulation;
- Improve risk analysis (FMEA, RCM, etc.) through the use of better system reliability models;
- Development of Monte Carlo simulation algorithms applied to reliability;
- In-depth study on the generation of censored data in order to be able to make better reliability simulations: development of algorithms for random censor data;
- Development of simulation algorithms of consecutive *k-out-of-n* linear systems that can be applied in the maintenance management of such systems or suitable preventive systems;
- Development of algorithms to simulate the calculation the components importance in complex systems;
- Development of the model and the algorithms for the use of the *EM - Expectation Maximization* method for the estimation of Weibull distribution parameters in the equipment reliability data;
- Application of the algorithms developed in a practical and real case;
- Development of a methodology that, using the company's information systems, applied advanced simulation tools;

- Contribute to an analysis that allows for better specification of the information system for maintenance management;

- Identify new patterns in the use of information systems for maintenance management.

1.3 Thesis Structure

The topic of the thesis covers areas with a very broad scope. Thus, more emphasis is given to the topics that are most relevant to achieve the objectives. Important concepts and terms will be explained when they are introduced in the text. However, the document was elaborated in a way that will allow any reader to understand the presented concepts and results.

This first chapter presents the relevance of the subject, the objectives of the dissertation, the influences and motivations and, finally, a description of the structure of the dissertation, as well as a summary in diagram form in order to become more visible and understandable.

In Chapter 2, there is a theoretical approach to the basic concepts of statistics, random variables and distribution functions used in the field of life time analysis. This chapter introduces the terminology and symbols that will be used throughout the thesis and defines the way in which the models and parameters are chosen from each of the statistical distributions.

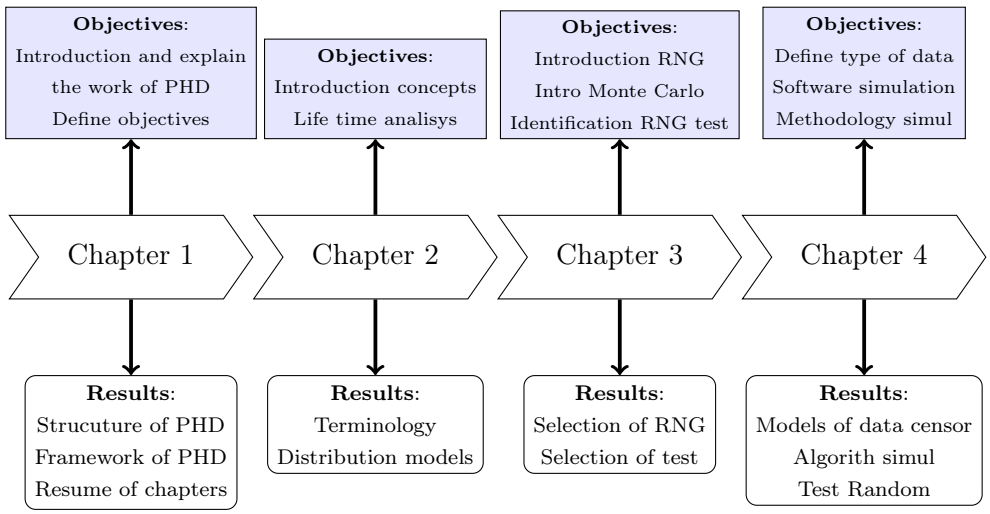


Figure 1.1: PhD structure - Chapters 1 to 4

Chapter 3 begins by providing a brief introduction to system simulation and modelling: models are real-world approximations. Then, the Monte Carlo method is described and explained as a technique, using random or pseudo-random numbers, for the solution of a model. It is now the most powerful and commonly used technique for analysing complex problems.

A summary of the state of the art in the generation of random numbers is the building block of a simulation study. The inverse transform technique used to sample, from the exponential, uniform, Weibull and triangular distribution, is referred, as are the other continuous distributions without a closed-form inverse transform technique. Additionally, the rejection method, one of the classic for generating random variables, is explained. Finally, the hypothesis test is provided below to identify whether the data obtained can be considered as random. The desirable properties of random numbers – uniformity and independence – need to be tested. The thesis will use the Kolmogorov-Smirnov test, Turning point test, Runs test, Rank test and Bartels test.

In Chapter 4, the censored data type is explained. Censoring data can arise for many sources and, depending on the sources, censoring may be of many different types. A definition and a proposal structure are provided for complete, truncated and censor data. A brief description of the three software programs used to develop and program the simulation data: *Matlab, Python and R* software, are all integrated in a suite of software facilities for data manipulation, calculation and graphical display. There is a sub-chapter to help define how to construct the algorithm for studies with censored data from different statistical distributions; at this point, a deep analysis and a huge number of simulations are carried out to compare the results and draw very consistent conclusions. This sub-chapter is an original proposal within the area of studies on reliability simulation. After this section, an algorithm is developed that provides the simulation for fixed right type I censored data, and like in the previous section, a deep study takes place. In this part and the next, an analysis methodology is designed, explained and applied. The last section is a study with the construction of an algorithm that carries out a simulation of type II censored data, that is, data that was censored at the end of n events or malfunctions.

In Chapter 5, the general formulation of censored data using *MLE - Maximum Likelihood Estimation* to achieve the unknown parameters of the statistical distributions include the mathematical expression for five statistical distributions complete which result in the estimation models for the various types of censored data (right, left, random, type I and type II). Algorithms are developed to compute a simulation of right censored type I data with Weibull distribution in *Python* software.

The next section describes the *EM - Expectation Maximization* algorithm. It is an iterative process that can be used to calculate the maximum likelihood estimators in situations with incomplete data. In this work, two software programs were used: *R and python* to solve and simulate the *EM method* for right censored data.

Chapter 6 begins with reliability and simulation of coherent systems that consider the structural relationship between a system and its components, that is, those relationships that are deterministic. The *RBD - Reliability Block Diagrams* theory is exposed as a success-oriented network describing the function of the system. It shows logical connections of (functioning) components needed to fulfil a specified system function. The theory and the reliability functions of the linear consecutive *k-out-of-n systems:F* and linear consecutive

k-out-of-n systems:G are derived and sub-section algorithms and simulation are carried out. Finally, the component importance measure is studied, and some of the indicators most applied in simulation model examples are chosen: *2-out-of-5*; Linear consecutive *2-out-of-5:G* and Linear consecutive *2-out-of-5:F*.

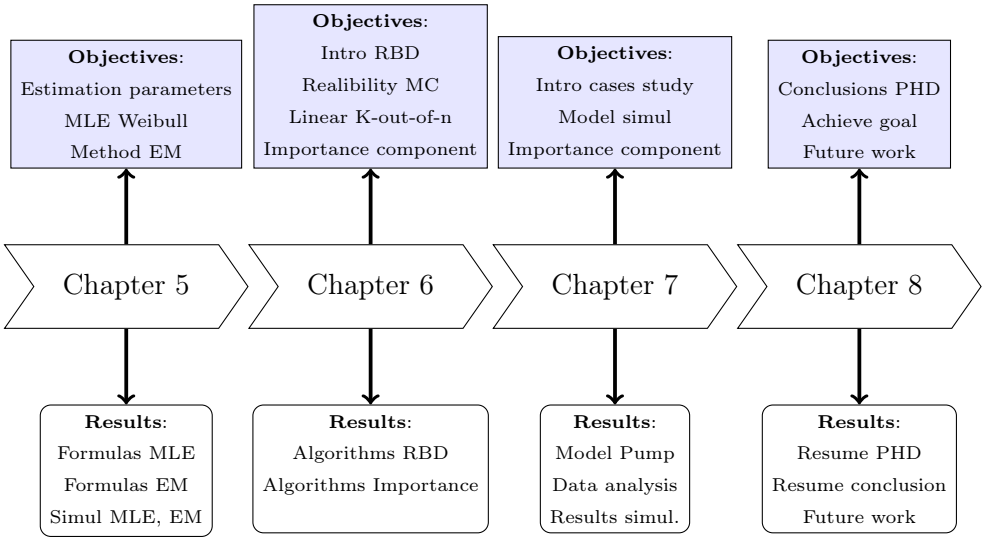


Figure 1.2: PhD structure - Chapters 5 to 8

The chapter 7 presents a case study on reliability analysis and simulation. Using the theory, methods and tools explained and developed in the last chapters, the case study is resolved and compared with simulation. The first part of the chapter begins to expose the methodological framework for modelling the system and applying the simulation in a real industrial company. After that, the equipment to be studied and its connections are described: the centrifugal pump in the petrochemical industry. The sub-chapter which follows is dedicated to data failure analysis of data reported for seven centrifugal pumps. Finally, the simulations carried out and the results obtained are analysed and compared. The end of the chapter presents the conclusions of the use of the methodology and how it can be applied in practical cases.

Chapter 8 presents the main conclusions and the future perspectives of the work. A new way of approaching simulation tools for maintenance management is proposed, with the intent of emphasizing the importance and the complexity of simulation within the industrial maintenance and in a complex and technological world.

Chapter 2

Life time analysis

The chapter presents the basic concepts for life time analysis, based on classic authors like Birolini (2017) and Ross (2017). It's an essential background for some of the chapters of the thesis. The focus is only on continuous distributions and have been chooses the most relevant distributions in reliability; there are some examples with data for demonstration of theory.

Life time analysis concerns to the life time of items or organic elements can both be applied for industrial or biological or another environments. Industrial equipment or product can't be in operation forever and can't always continue to operate in the original state; the same happen to biological life being. The change in conditions or the breakdown can be termed as a failure and in the world of biological life a disease or death of the life organism.

The life time of the item, or organism life is the timespan from the beginning of operation or birth, until the failure or death. The study of lifetime from items (industrial) or products (domestic) is called reliability analysis and the lifetime of biological units is termed as survival analysis.

A failure or breakdown of an item, equipment or product does not necessarily means the end of the actual life time, they can be repaired or refurbished. Assuming the failure of equipment when the item is not able to do the function that are required, for example, a pump that can work but can't give the necessary power to pump the liquid or give the right pressure (or elevation) that is necessary to the fluid; the pump is not in breakdown situation, but can't do the function that is designed. Otherwise, for a biological unit a death can be a failure or an occurrence of a certain disease.

The general definition of lifetime analysis considers that the time could be not either real time or clock time. An example is the landing gear from air-plane when the lifetime is defined by the number of landings. The lifetime data is a type of data that measures a variable parameter , until the occurrence of an event that is in study, can be failure, death, abnormal function, disease, etc... In general, a failure can be considered as an event.

The lifetime analysis is very important in many areas of knowledge in the modern society. In analysis, the first interest is to know how variable t fit the data to an underlying distribution. We assume that the data are collected, organized and arranged, so the first step is very important but is outside of the scope of the thesis. After choosing the appropriate distribution model that fits the data (using hypothesis test to select or confirm the best model distribution), it's possible to infer about the different characteristic of the data of the item that is in study. For example, it can be estimated certain parameters related of lifetime variable, such as *MTBF - Mean Time Between Failure* or the probability for the item to be in operation in the future. There are two different models to the estimation and inference: parametric and non-parametric. Both are used in practice and in research.

2.1 Basic concepts and terminology

Introduced by *kolmogorov (2013)*, the *probability space* $[\Omega, \mathbb{F}, P_r]$ is a mathematical model of an experiment with random outcome. Ω is the *sample space*, $\mathbf{F}$ the *event field*, and P_r the *probability* of each element of $\mathbf{F}$. The *sample space* Ω is a set containing all possible outcomes of the experiment in study.

Consider, for example, the *sample space* $\Omega = \{1, 2, 3, 4, 5, 6\}$ that represent the experiment of a single throw of a die, and other one, $\Omega = [0, \infty)$ in the case of failure time of an equipment or component. The elements of Ω are called *elementary events* and are represent by ω. The outcome of the experiment is a subset Γ of Ω; combinations of statement become equivalent to operations with subsets of Ω. If the sample space Ω is finite or countable, a probability can be assigned to every subset of Ω and, if Ω is continuous, restrictions are necessary.

The *event field* $\mathbb{F}$ is thus a system of subsets of Ω which a probability can be assigned. Such field has the following properties:

1. Ω is an element of $\mathbb{F}$.
2. If A is an element of $\mathbb{F}$, its complement $\bar{\mathrm{A}}$ is also an element of $\mathbb{F}$.
3. If $A_1, A_2, ...$ are elements of $\mathbb{F}$, the countable union $A_1 \cup A_2 \cup ...$ is also an element of $\mathbb{F}$.

2.1.1 Random variables

When performing an experience the objective is concerned about the value of a numerical quantity determined by the output of experiment. These values of interest that are determined by the output of the experience are known as random variables.

The cumulative distribution function *(CDF)*, or simply the distribution function, F of the random variable X is defined for any real number x by:

$$F(x) = P\{X \leq x\}$$

A random variable that can take a countable number of possible values is said to be discrete. For a discrete random variable X we define its probability mass function *p(x)* by

$$p(x) = P\{X = x\}$$

If X is an discrete random variable that takes one of the possible values $x_1, x_2, ...$, then, since X must take one of these values we have:

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Example 2.1.1:
Suppose that X takes one of the values *0,1 or 2*. If

$$p(0) = \frac{1}{3}, \qquad p(1) = \frac{1}{4}$$

then, since $p(0) + p(1) + p(2) = 1$, it follows that $p(2) = \frac{5}{12}$

A discrete random variable is a countable set of possible values and often have to consider random variables whose set of possible values is an interval. The random variable x is continuous random variable if there is a non-negative function *f(x)* defined for all real number x and having the property that for any set D of real numbers

$$P\{X \in D\} = \int_D f(x)dx \tag{2.1}$$

The function f is named the *PDF - probability density function* of the random variable X. The PDF satisfies the following conditions:

1. $f(x) \geq 0$ for all x in D.
2. $\int_D f(x)dx = 1$.
3. $f(x) = 0$ if x is not in D.

For a continuous random variable X, the probability that is in the interval [a,b] is given by

$$P(a \leq X \leq b) = \int_a^b f(x)dx \tag{2.2}$$

The graphical interpretation of (2.2) is shown in figure 2.1. The shaded area represents the probability that X is in the interval [0.5,2]

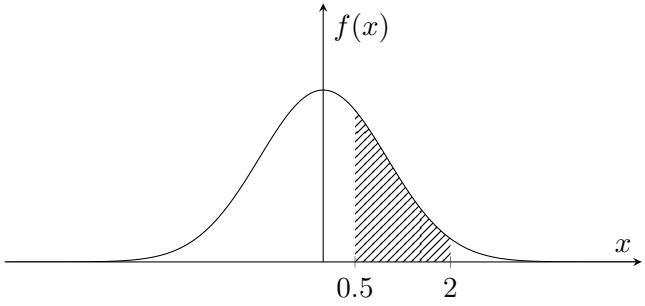


Figure 2.1: *f(x)* - probability density function of random variable x

The relationship between the CDF $F(\cdot)$ and the PDF $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a)\} = \int_{-\infty}^{a} f(x)dx$$

Differentiating both sides of the equation:

$$\frac{d}{da}F(a) = f(a)$$

The density function results as the derivative of cumulative distribution function. A intuitive interpretation of the density function may be obtained from equation (2.1) as follow:

$$P\left\{a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}\right\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)dx \approx \varepsilon \cdot f(a)$$

When ε is small the probability that X will be in an interval of length ε around the point a is approximately $\varepsilon \cdot f(a)$. With this result the conclusion is that *f(a)* is a measure of how likely it is that the random variable will be near a.

The cumulative distribution function denoted by *F(x)*, measures the probability that the random variable X assume a value less than or equal to x, that is:

$$F(x) = P(X \leq x).$$

If X is discrete, then

$$F(x) = \sum_{all\ x_i \leq x} p(x_i)$$

If X is continuous, then

$$F(x) = \int_{-\infty}^{x} f(x)dx \tag{2.3}$$

Some properties of the CDF are:

1. F is a non-decreasing function. If $a < b$, then $F(a) \leq F(b)$.
2. $\lim_{x \to -\infty} F(x) = 0$.
3. $\lim_{x \to \infty} F(x) = 1$.

The probability analysis from X can be answered in terms of CFD, for example:

$$P(a < X \leq b) = F(b) - F(a) \quad \text{for all} \quad a < b$$

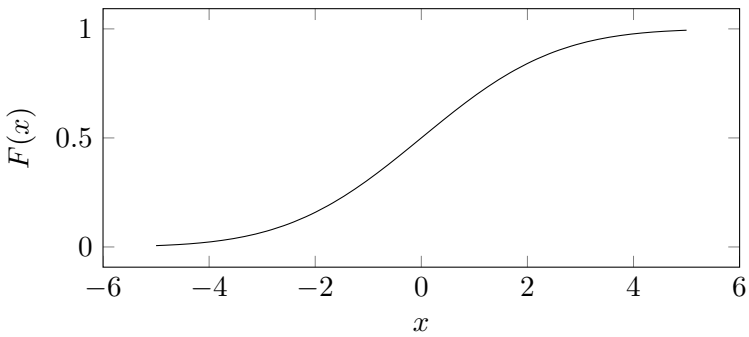


Figure 2.2: *F(x)* - cumulative distribution function of random variable x

2.1.2 Expectation

In probability theory an important concept is that of the expectation of a random variable. If X is a random variable, the expected value of X, denoted by *E(x)*, for discrete and continuous variable is defined as follows:

$$E(x) = \sum_{\text{all } i} x_i p(x_i) \qquad \text{if X is discrete}$$

and

$$E(x) = \int_{-\infty}^{\infty} x f(x) \qquad \text{if X is continuous}$$

The expected value E(x) of a random variable X is also referred to as the mean μ, or the first moment of X. The quantity $E(X^n), n \geq 1$, is called the nth moment of X.

The variance of a random variable X, denoted by *V(X)* or *Var (X)* or σ^2, is defined by

$$V(X) = E[(X - E[X])^2].$$

A useful identity in computing *V(X)* is given by

$$V(X) = E(X^2) - [E(X)]^2.$$

The mean *E(X)* is a measure of the central tendency of a random variable.

The variance of X measures the expected square of the random variable form its expected value. Thus, the variance, *V(X)*, is a measure of the spread or variation of the possible values of X around the mean *E(X)*.

The standard deviation, σ, is defined to be the square root of the variance, σ^2. The mean, *E(X)*, and the standard deviation, $\sigma = \sqrt{V(X)}$, are expressed in the same units.

2.1.3 Reliability statistics

Let $T_1, T_2, ..., T_n$ be a sample of size n of an continuous random variable T. For each $i = 1, 2, ..., n$, the observation t_i represents the lifetime of the individual i of the random variable T_i. The sample is independent and identically distributed and defined on the probability space $(\Omega, \mathbb{F}; P)$.

Let $f(t)$ be the probability density function of T and because the equipment in reliability theory implies $t > 0$, and t denotes the failure time (operation time free of failure) of an item, distributed according to its cumulative distribution function F(t):

$$F(t) = Pr\{T \leq t\} \quad \text{with F(0)=0}$$

The reliability function *R(t)* gives the probability that the equipment or component will operate without failure in $(0, t)$; thus,

$$R(t) = Pr\{T > t\} = 1 - F(t), \quad t > 0,$$

or survival function usually adopted for lifetime analysis

$$S(t) = Pr\{T > t\} = 1 - F(t), \quad t > 0.$$

The survival function gives the probability that a subject or item survive past time t. As t ranges from 0 to ∞, the survival or reliability function has the following properties:

- It is non-increasing.
- At time *t=0*, *S(t)=1*. The probability of surviving at time 0 is 1.
- At time $t = \infty, S(t) = S(\infty) = 0$. As time goes to infinity, the survival curve goes to 0.

In theory the difference from the survival or reliability function is smooth. In practice, it observes events on a discrete time scale (days, weeks, etc...). The failure rate $\lambda(t)$ of an item exhibiting a continuous time to failure T is defined as the conditional probability of a failure in the interval t to $t + \delta t$ given that there was no previous failure until t

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\delta t} Pr\{t < T \leqslant t + \Delta t \mid T > t\}.$$

If T and t are statistically independent, then through the following calculations we obtain

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\delta t} * \frac{Pr\{(t < T \leqslant t + \Delta t) \cap (T > t)\}}{Pr\{T > t\}} = \lim_{\Delta t \to 0} \frac{1}{\delta t} * \frac{Pr\{t < T \leqslant t + \Delta t\}}{Pr\{T > t\}},$$

and thus, assuming $F(t)$ derivable,

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = -\frac{dR(t)/dt}{R(t)} \quad t > 0. \tag{2.4}$$

and

$$-\frac{dR(t, x_0)/dt}{R(t, x_0)} = \lambda(t, x_0) = \lambda(t + x_0) \tag{2.5}$$

The failure rate $\lambda(t)$ is defined as conditional density for failure in $(t, t + \delta t)$ given that the item was new at *t=0* and has not failed in $(0, t)$ and is distinguished from density *f(t)*, as unconditional density for failure in $(t, t + \delta t)$ given only that the item was new at *t=0* (assumed with *F(0)=0*).

The failure rate $\lambda(t)$ applies in particular to non-repairable items. However, considering equation 2.4, it can also be defined for repairable items which are as good-as-new after repair (renewal), taking instead of t the variable x starting by *x=0* at each renewal.

If a repairable item cannot be restored to be as-good-as-new after repair, it's used failure intensity *z(t)*.

Considering the equation 2.5:

$$\lambda(t, x_0) = -\frac{dR(t, x_0)/dt}{R(t, x_0)}$$

with the following development:

$$\lambda(t)dt = -\frac{1}{R(t)} dR(t) \Leftrightarrow$$

$$\Leftrightarrow \int_0^t -\frac{1}{R(t)} dR(t) = \int_0^t \lambda(t)dt \Leftrightarrow$$

$$\Leftrightarrow [\ln R(0) - \ln R(t)] + c = \int_0^t \lambda(t)dt \Leftrightarrow$$

$$\Leftrightarrow -\ln R(t) + c = \int_0^t \lambda(t)dt \Leftrightarrow$$

$$\Leftrightarrow R(t) = \exp^{-\int_0^t \lambda(x)dx}$$

and yields

$$Pr\{T > t\} = R(t) = e^{-\int_0^t \lambda(x)dx}.$$

Similarly, instead of failure rate the survival analysis considers the hazard function, *h(t)*, that is the instantaneous rate at which events occur, given no previous events.

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t < T \leqslant t + \Delta t \mid T > t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{2.6}$$

The cumulative hazard $H(t) = \int_0^t h(u)du$ describes the accumulates risk up to time t and S(t) as survival function. It's possible derive the other two functions, if we know any one of the function *S(t), H(t), h(t)*:

$$h(t) = -\frac{\partial \log(S(t))}{\partial t}$$
$$H(t) = -\log(S(t))$$
$$S(t) = exp(-H(t))$$

2.2 Life time distribution models

Not all statistical distributions are used as models for lifetime data. There are certain distributions, which are useful in depicting the lifetime data due to some desirable properties that they have. Here, we briefly describe some statistical distributions that are commonly used for modelling lifetime data. For more details on these distributions, one may refer to O'Connor and Kleyner (2012), Bradley (2016).

2.2.1 Uniform distribution

A random variable X is said to be uniform distributed over the interval $(a, b), \quad a < b$, if its probability density function is given by:

$$f(x) = \begin{cases} \frac{1}{b-a} & if \quad a < x < b, \\ 0 & otherwise \end{cases}$$

In other words, X is uniformly distributed over *(a,b)* if it have its mass on that interval and it is equally likely to be "near" any point on that interval.

The mean and variance of a uniform *(a,b)* random variable are obtained as follows:

$$E[X] = \frac{1}{b-a}\int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

$$\sigma^2(X) = \frac{1}{b-a}\int_a^b x^2 dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + b^2 + ab}{3}$$

$$Var(X) = \frac{1}{3}(a^2 + b^2 + ab) - \frac{1}{4}(a^2 + b^2 + 2ab) = \frac{1}{12}(b-a)^2$$

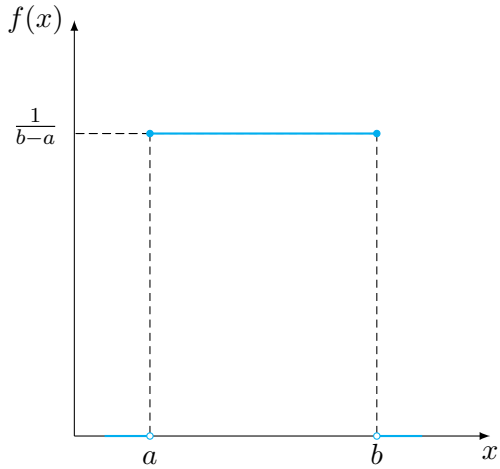


Figure 2.3: *f(x)* - probability density function for Uniform distribution

The expected value is the midpoint of the interval (a,b). The distribution function of X is given for $a < x < b$, by:

$$F(x) = P\{X \leq x\} = \int_a^b (b-a)^{-1} dx = \frac{x-a}{b-a}$$

2.2.2 Exponential distribution

The exponential distribution is characterized when the hazard rate is constant. It is considered a homogeneous poisson process.

A continuous random variable having probability density function:

$$f(x|\lambda) = \begin{cases} 0 & (x \leq 0), \\ \lambda e^{-\lambda x} & (x > 0), \end{cases}$$

for some $\lambda > 0$ is said to be an exponential random variable with parameter λ and symbolically express, $X \sim Ex(\lambda)$.

Its cumulative distribution is given by:

$$F(x) = \int_0^x \lambda e^{-\lambda x} = 1 - e^{-\lambda x}, \quad 0 < x < \infty$$

It is easy to verify that the expected value and variance of such a random variable are as follow:

$$E(X) = \frac{1}{\lambda} \quad and \quad Var(X) = \frac{1}{\lambda^2}$$

This is an important distribution in reliability field, as it has the central limit theorem relationship to statistics as the normal distribution has to non-life statistics. It describes the constant hazard rate behaviour. As the hazard rate is often a function of time, we will denote the independent variable by t instead of x. The constant hazard rate is denoted

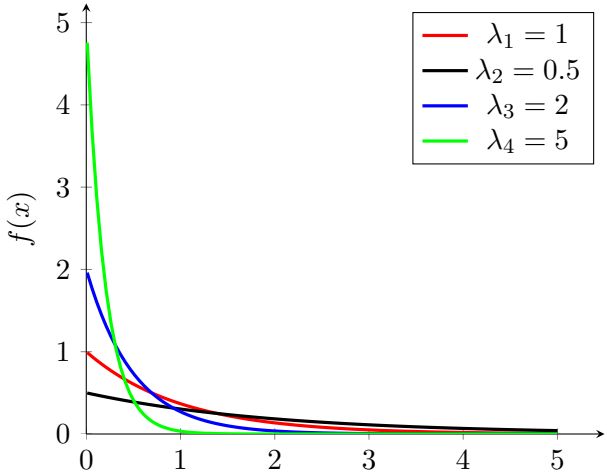


Figure 2.4: *f(x)* - probability density function for Exponential distribution

by λ. The exponential distribution show that the mean time of waiting (in a queue for example) is proportional inverse from the frequency of the events; for example, if mean have 60 events per hour, the mean time of waiting will be 1 minute.

The exponential distribution doesn't have mode, because the density function is decreasing with x and the domain is open. The median is $\mu_e = (ln2)/\lambda$.

The exponential distribution have an important property, that is called "memoryless property" and can formalize in that way:

$$T \sim Ex(\lambda) \Rightarrow P(T > t + h | T > t) = P(T > h). \tag{2.7}$$

To understand this property let take T the variable that represents the lifetime of some unit, thus equation (2.7) is a statement of fact that the distribution of the remaining life of an item of age h does not depend on h. That is, it is not necessary to remember the age of the unit to know its distribution of remaining life. Equation (2.7) is equivalent to

$$P\{\text{additional life of an item of age h exceeds t}\} = P\{T > h + t | T > h\}$$

In this case, the item is not ageing, and if, for example, the item survives until ten years, the survival distribution is the same that the item have in the beginning. In practical applications, sometimes the exponential distribution is not used because don't have in account the ageing and wear-out.

Another useful property of exponential random variable is that they remain exponential when multiplied by a positive constant. To see this suppose that T is exponential with parameter λ, and let c be a positive number. Then

$$P\{cT \leq t\} = P\left\{T \leq \frac{t}{c}\right\} = 1 - e^{-\lambda \frac{t}{c}}$$

which shows that cT is exponential with parameter λ/c.

2.2.3 Gamma distribution

The gamma distribution represents the sum of n exponentially distributed random variables. The gamma distribution is a life distribution model that gives a good fit to some applications of failure data, and it's very flexible distribution and it describes the situation when partial failure can exist.

For a random variable, X, and symbolically write, $X \sim G(\alpha, \lambda)$, the probability density function is given by,

$$f(x, \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{(\alpha-1)} e^{-\lambda x}, \quad x > 0 \quad \alpha, \lambda > 0$$

Where λ is the failure rate (complete failures) is calculate from $\lambda(t) = f(t)/(1 - F(t))$. $\lambda(t)$ is constant (time-independent) for $\alpha = 1$, *monotonically decreasing* for $\alpha < 1$ and *monotonically increasing* for $\alpha > 1$. α is the number of partial failure per complete failure or events to generate a failure.

$\Gamma(\alpha)$ is the *gamma function*:

$$\Gamma(a) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

When (a-1) is a positive integer, $\Gamma(a) = (a-1)!$ is a situation case of the partial failure.

The exponential distribution is a special case of the gamma distribution when $\alpha = 1$, that is,

$$f(x) = \lambda e^{-\lambda x}$$

However, in contrast to the Weibull distribution, $\lambda(t)$ always converges to λ for $t \to \infty$. A gamma distribution with $\alpha < 1$ mixed with a three-parameter Weibull distribution can be used as an approximation to the distribution function for an item with failure rate as the *bathtub curve*.

The mean and the variance are given by

$$E[X] = \frac{\alpha}{\lambda} \quad and \quad Var[X] = \frac{\alpha}{\lambda^2}$$

It follows that the sum of two independent gamma-distributed random variables with parameters λ, α_1 and λ, α_2 has a gamma distribution with parameters $\lambda, \alpha_1 + \alpha_2$

The figure 2.5 shows the density functions of the gamma distribution for different parameter values, which allows to verify that this distribution covers very different situations.

When the α parameter is integer, the gamma distribution is known as the Erlang distribution, and, when $\alpha = 1$, results in the particular case of the exponential distribution.

2.2.4 Weibull distribution

The Weibull distribution is the most popular statistical distribution used in reliability engineering. It can be used to fit many life distributions and it has the great advantage in

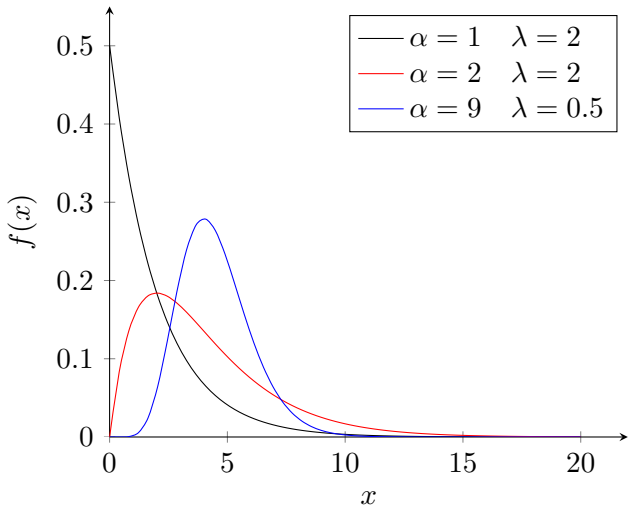


Figure 2.5: *f(x)* - probability density function for Gamma distribution

reliability field because adjusts well to the reliability data.

The density function is given by

$$f(x, \eta, \beta) = \frac{\beta}{\eta^\beta} t^{\beta-1} e^{-(\frac{t}{\eta})^\beta} \; with \; t \in \mathbf{R}^+$$

And the corresponding reliability function is

$$R(t) = e^{-\left(\frac{t}{\eta}\right)^\beta}$$

The hazard rate is

$$\lambda(t) = \frac{\beta}{\eta^\beta} t^{\beta-1}$$

The shape parameter β is a non-dimensional parameter and reflect the type of failure mode, such as infant mortality ($\beta < 1$), random or exponential ($\beta = 1$), or wear-out ($\beta > 1$).

The scale parameter η has the same unit as t, or *characteristic life* - it is life at which 63.2% of the population will have failed and is a function of the $MTTF$ – Mean Time To Failure with a general relationship given by the following equation:

$$MTTF = E(T) = \mu = \eta\Gamma(1 + 1/\beta)$$

The other statistical characteristic that should be calculated during a typical Weibull analysis procedure being applied to analysis the failure data is the variance given by the following equation:

$$Var[t] = \eta^2(\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta))$$

An analytical expression for the *Laplace transform* of the Weibull distribution function does not exist.

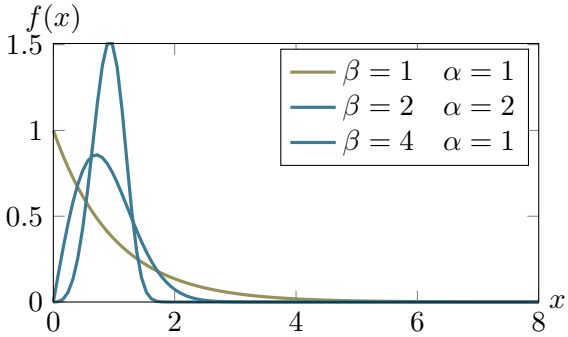


Figure 2.6: *f(x)* - probability density function for Weibull distribution

2.2.5 Normal distribution

The most widely used statistical distribution in theory and practice is the function named as the normal or Gaussian. From the point of view of practical experiences its has been verified that a lot of observable variables of a certain population are represented by a normal distribution.

The normal distribution has important properties that make it very attractive for practical applications and for theoretical studies such as:

- Description of physical phenomena
- Biometric variables (weight, height, length, etc.)
- Variables linked to production and quality
- In many limit situations, the other distributions converge to the distribution Normal, as the central limit theorem explains.

A random variable X has a standard distribution with parameters μ and σ^2 with a density function of the form:

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

It is written symbolically, $X \sim N(\mu, \sigma^2)$

The parameters of the normal distribution are represented by μ and σ^2 because they correspond, respectively, to the mean and variance of the random variable. Some authors prefer to employ the σ (standard deviation) parameter.

The cumulative distribution function is defined by the integral

$$F(x|\mu,\sigma^2) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\}} dt,$$

for which no analytical solution is known. The values of the distribution function must be calculated using numerical analysis methods.

An important reason for the wide applicability of the normal distribution is the fact that whenever several random variables are added together, the resulting sum tends to normal regardless of the distribution of the variables being added. This is known as the *central limit theorem*. It justifies the use of the normal distribution in many engineering applications, including maintenance. The normal distribution is close fit to most quality control applications and some reliability observations, such as the lives of items subject to wear-out failures.

The N (0,1) distribution is usually referred to as standardized normal distribution. This particular case, where $\mu = 0$ and $\sigma^2 = 1$, is very important and is especially useful once that their values are tabulated.

If $X \sim N(\mu, \sigma^2)$ the variable $Z = \frac{X-\mu}{\sigma}$ has distribution N(0,1).

If we consider the change of variable defined by the standardized variable

$$Z = \frac{X - \mu}{\sigma}$$

the density function and the distribution function are, respectively:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}},$$
$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{z^2}{2}} dt.$$

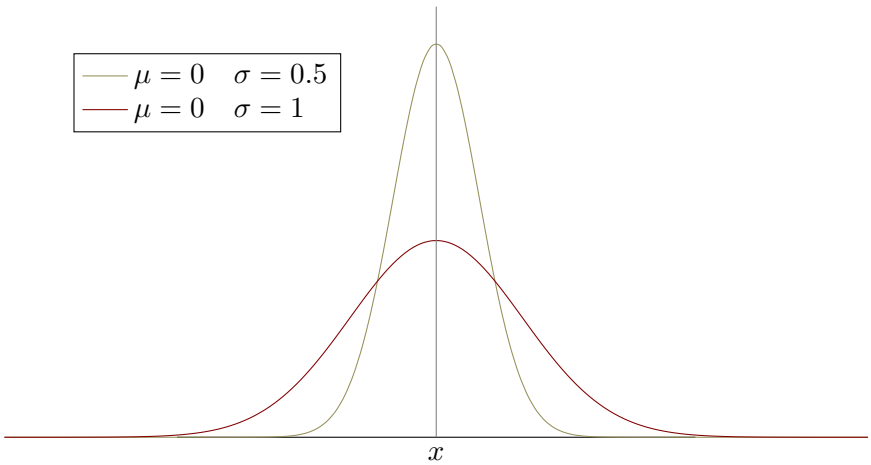


Figure 2.7: *f(x)* - probability density function for Normal distribution

The figure 2.7 represents the density functions of the normal distribution for some values of μ and σ^2. From the analysis of the expression $f(x|\mu, \sigma^2)$ we can deduce that the normal density function is symmetric with respect to the straight line $x = \mu$.

The normal distribution *N(0,1)* is largely tabulated. The first aspect to be taken into account when referring to the tables is that they refer to $\Phi(z)$, and therefore allows to calculate for each z the respective ordinate of the standardized normal distribution function. For example, if $z = 0.75$, results $\Phi(0.75) = P(Z < 0.75) = P(Z \leq 0.75) = 0.77$. The Normal table generally refers only to non-negative values of u.

However, the symmetry (with respect to 0) of the normal distribution ensures that $\Phi(z) = 1 - \Phi(-z)$, which allows to immediately determine values of $\Phi(z)$ to negative values. For example, if $z = -0.3$, results $\Phi(-0.3) = 1 - \Phi(0.3) = 1 - 0.617 = 0.382$.

Example 2.2.1:
The annual precipitation (in mm) in the Viseu district is well modelled by a normal distribution with $\mu = 900$ mm and $\sigma = 70$ mm. Suppose that it was intended to calculate the probability that the annual precipitation would be between 800 and 1000 mm.

$$P(800 < X < 1000) = \int_{800}^{1000} f(x)dx = F(1000) - F(800) = \Phi\left(\frac{1000-900}{70}\right) - \Phi\left(\frac{800-900}{70}\right)$$
$$= \Phi(1.42) - \Phi(-1.42) = 0.922 - 0.078 = 0.83$$

Suppose now that it was intended to determine the probability that the annual precipitation would be less than 850 mm.

$$P(X < 850) = F(850) = \int_{-\infty}^{850} f(x)dx = \Phi\left(\frac{850-900}{70}\right) =$$
$$= \Phi(-0.71) = 1 - \Phi(0.71) = 1 - 0.76 = 0.24$$

2.2.6 Log-normal distribution

A random variable is log-normally distributed if the logarithm of the random variable is normally distributed. The log-normal is a versatile distribution and often a better fit to reliability data, such as for populations with wear-out characteristics. The log-normal distribution have an advantage to the normal distribution in reliability applications because it doesn't extend below zero. It's used to model usage data, such as repair time of maintenance, pumps and motors in hours per year, mechanical devices, etc...

The log-normal PDF is:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(lnx-\mu)^2}.$$

The log-normal is the normal distribution with $ln\ x$ as the variable. The mean and the variance of the log-normal distribution are given by:

$$E[X] = exp(\mu + \frac{\sigma^2}{2})$$

$$Var[X] = [exp(2\mu + 2\sigma^2) - exp(2\mu + \sigma^2)]^{1/2}$$

where μ and σ are the mean and the standard deviation of the *ln* data. When $\mu \gg \sigma$, the lognormal distribution approximates to the normal distribution.

The lognormal distribution describe reliability of items in which the hazard rate increases from $x{=}0$ to a maximum and then decreases.

2.2.7 Triangular distribution

A random variable X has a triangular distribution if its pdf is given by

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & a \leq x \leq b, \\ \frac{2(c-x)}{(c-b)(c-a)} & b \leq x \leq c, \\ 0, & elsewhere \end{cases}$$

where $a \leq b \leq c$. The mode occurs at $x{=}b$. A triangular pdf is shown in figure 2.8. The parameters *(a,b,c)* can be related to other measures, such as the mean and the mode, as follows:

$$E(X) = \frac{a+b+c}{3} \tag{2.8}$$

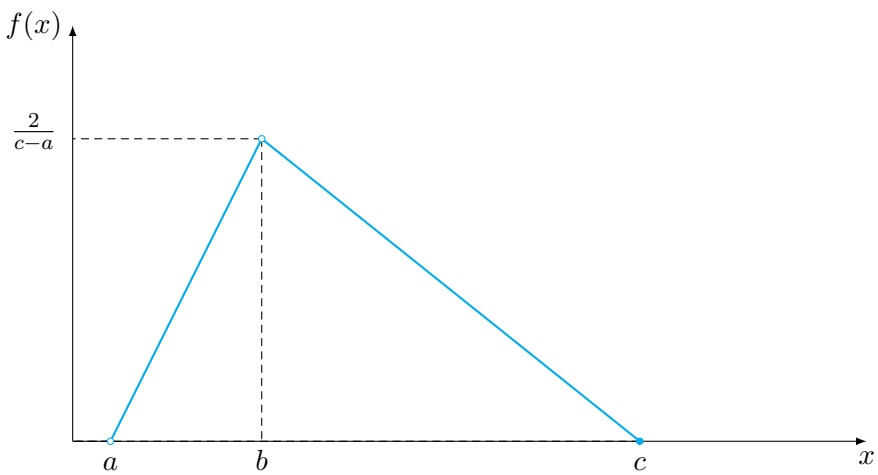


Figure 2.8: *f(x)* - probability density function for Triangular distribution

From equation (2.8) the mode can be determined as

$$Mode = b = 3E(x) - (a + c).$$

Since $a \leq b \leq c$, it follows that

$$\frac{2a+c}{3} \leq E(X) \leq \frac{a+2c}{3}.$$

The mode is used more often than the mean to characterize the triangular distribution. As shown in figure (2.8) its height is $2/(c-a)$ above the x axis. The CDF for the triangular distribution is given by:

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{(x-a)^2}{(b-a)(c-a)} & a < x \leq b, \\ 1 - \frac{(c-x)^2}{(c-b)(c-a)} & b < x \leq c, \\ 1, & x > c. \end{cases}$$

2.3 Statistical confidence

Statistical confidence studies are used in making an assertion about a population given data from a sample and to determined the confidence interval which includes the true value. The confidence interval is the interval between the *upper* and *lower* confidence limits, if the experiment is replicated many times. the higher the number of sample more will be our intuitive confidence that the estimate of the population parameter will be close to the true value.

Statistical confidence is different from engineering confidence; statistical confidence takes no account of engineering or process knowledge. The confidence interval must always be interpreted in the light of engineering knowledge, which might serve to increase or decrease our engineering confidence.

If the population value x follows a normal distribution, it can be shown that the means, $\bar{x}$, of samples drawn from it are also normally distributed, with variance σ^2/n. The Standard variance of the sample mean is also called the *standard error of the estimate.*

If x is not normally distributed, provided that n is large ($n > 30$), $\bar{x}$ will tend to a normal distribution. If the distribution of x is not excessively skewed (and is unimodal) the normal approximation for $\bar{x}$ at values of n as small as *6 or 7* may be acceptable. These results are derived from the central limit theorem. They are of great importance in deriving confidence limits on population parameters, based on sample data. In reliability field, it is not usually necessary to derive exact confidence limits and therefore the approximate methods described are quite adequate.

Chapter 3

Simulation, random number generator and statistical test

The use of methodologies that help to eliminate uncertainty in the business decision-making process is of enormous importance in modern business. The advent of modelling and simulation provides itself as an innovative and useful support tool for improving process and business management.

This chapter presents a brief theoretical introduction to the concepts of modelling and simulation, followed by the description and explanation of *RNG - random number generating* and the last part, the introduction of statistical test of hypothesis for randomness.

There are many reasons guiding the search for knowledge and the need to understand what surrounds us. One of the first method of learning about the nature was experimentation, and for thousands of years remains one of the main ways of solving problems. In the experimentation there are two entities, the object studied and the framework of conditions that defines how the experiment takes place.

However, sometimes the problems to be solved are so complex that experimentation may prove to be an inadequate methodology due to ethical, risk or cost issues. In other cases, the solution is simply impossible to put into practice. Science has been able to find ways to address these obstacles, one of which is to make abstraction from the problem and think of it using a conceptual model.

3.1 Simulation and modelation

Solving problems is in human nature. The techniques used in this type of problem solving are analytical in the sense that they are symbolic and based on reasoning, providing generic solutions to the problems faced. An abstraction of the entity is made and the resulting model is then used to evaluate the system. This abstraction is made with loss of information (simplifications), but it represents the behaviour of the entities, allows analysis and is sufficient to demonstrate the properties of the proposed model. It is essential that the

results of the model are in agreement with those observed in the entity, in which case they can be considered valid.

Starting in the 1940s, computers provided alternative ways of dealing with these matters, since they are well equipped to apply approximation techniques, eliminate human computation errors, and solve problems at a higher speed. Thus, since the advent of computing, the numerical models have been converted into computational solutions (also called computational simulations). A computational simulation cycle can also be developed with the difference that the model is executed by specialized devices. Verification continues to be necessary, as precision limitations can cause erroneous results and diverge from expected solutions.

The simulation enables experiments with virtual environments, advancing the level of analysis of natural and artificial applications to levels unprecedented in scientific history, and allowing the design and analysis of complex applications. The simulation also provided cost-effective and risk-free training solutions when compared to experimentation. The simulation is a mature and economical problem-solving approach, which evolution fast with increasing computational power available.

According to Banks et al. (2013), at the heart of the *M&S - Modelling and Simulation* discipline is the notion that "models are real-world approximations". The first step in *M&S* is the creation of a model that represents an event or system, which can then be modified and the simulation allows observation of its behaviour. After performing simulations on the model, the analysis takes place to draw conclusions, verify and validate the relevant research and make recommendations. Visualization provides a way of interacting with the models, and is a form of data representation. The foundations of *M&S* are based on these four concepts:

1. Modeling
2. Simulation
3. Preview
4. Analysis

In *M&S* a system is a collection from different elements that together produce results and is the subject or thing to be investigated, and the object of the model development. A quantitative evaluation of the system is important to the modeller, noting how the system reacts to multiple inputs in different environments using objective evaluation criteria.

A model is a physical, mathematical, or logical representation of a system, entity, phenomenon or process. Its function is to give the best representation of real events and/or things of a real system, mainly because these can be difficult, if not impossible, to investigate. Essential systems that can not be removed from operation or theoretical systems without physical parts where to conduct experiments have to developed models

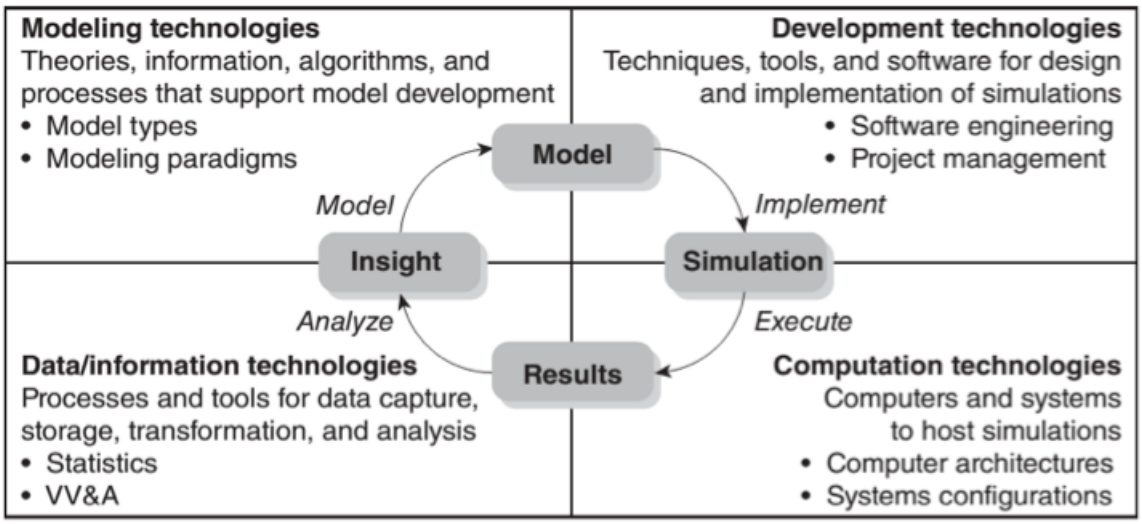


Figure 3.1: The Modelling and Simulation ($M\&S$) process (Sokolowski and Banks 2010)

and the study is carried out on them. The development of a model occurs by abstracting from reality using a description of the system. Only the characteristics of the system that affect its performance need to be represented in the model; the exhaustive inclusion of all the characteristics of the real system would imply high costs, time consuming, and excessive complexity being perhaps impossible. The model should be simple as possible, but have to represents the system. It is a challenge for the modeller to decide which system features should be included in the model.

A model can be physical or a set of mathematical equations or logical statements that describe the behaviour of a system (theoretical/notional systems). Mathematical models can result in analytical solutions (mathematical proof) if the complexity is low, or numerical solutions in other cases (with associated degree of uncertainty). The process of solving numerically a problem in this context is referred to as computer simulation, the definition of simulation is more complex than the model definition.

Therefore, a simulation can be understood as a method that describes the behaviour of a system using a mathematical or symbolic model, due to the fact that interaction with a real system may not be possible due to inaccessibility, risk, cost or unacceptable interaction with the system and even the possibility that the system does not even exist. A single implementation of the simulation is referred to as a test; a series of simulation trials is an exercise, and may occur as part of a learning process.

Another important concept is $M\&S$ it refers to the process of developing a model of a system and then executing simulations on that model in order to collect data related to its performance. The resulting data can serve as a basis for management, training and technical decisions, and when systems of large size and complexity are concerned, simulation may be the only viable tool for its analysis. $M\&S$ begins with the development of a computer simulation model based on a real or theoretical system and then runs the

model on a computer and analyses the resulting data.

According to Sokolowski and Banks (2010) the *M&S* development cycle through four phases, each requiring a different set of associated technologies, as shown in figure 3.1. It is only after a number of iterations that the process will reach its optimum efficiency. The first iteration is useful for providing information and data for model optimization. Good practices recommend repetition of the process until the members of the participating teams are satisfied with the results obtained that produce a reliable representation of the system studied.

3.2 Monte Carlo Simulation

The reliability of the system measures its ability to operate when a demand appears in industrial or service environment. However, after a long development in reliability science it still continues to be difficult to the maintenance manager to do some good work to predict and optimize the best interval of time for maintenance and the time to failure of equipment. That difficulty have contributed to make distant the relation between the academic or scientific world of reliability and the professional staff of maintenance which shows us that the actual models of reliability are not so good enough.

This section discusses the reliability methods and the use of mathematical methods, in particular the Monte Carlo method, as tools that can support and improve the engineering and reliability of systems. To study the characteristics, accuracy and related problems of different Monte Carlo techniques some typical examples have been selected, analysed and discussed.

Modern science uses the rules and methods to enable prediction. The predictions calculated in physics are very accurate, but to other science like social sciences or reliability the predictions are not so accurate and normally very difficult to calculate. Systems usually operate according to coherent laws but they are affected by internal and external random events and interactions. Their behaviour is inherently stochastic, not deterministic one in a very controlled and defined situation, so, its necessary to use mathematical and statistical methods, and, in particular, the Monte Carlo method in the prediction process to have some accuracy.

One thing that defines some deep difference between our time and the ancient empires, is the modern concept of reliability and availability. In the old empires the things were done to work or serve like bridges, houses and utilities. The main problem was how to produced or to developed the project, that some people had desired (in principle the King). Today the productions of things are normal and effortlessly, the design of the product is done taking into account how much time the product will work until the first failure or time of life cycle of the product. Reliability is a concept that most of the occidental people have in their mind, since the designer engineering until the consumer with low qualification. It's

usual that people ask how much it costs the maintenance of the car by year or by a long period of time, and how many years they can have that car with no problem.

The Monte Carlo method was considered to be a technique, using random numbers, for solution of a model and is now the most powerful and commonly used technique for analysing complex problems. It's application can be found in many fields of the science. Today it is recognized as playing an important role in system reliability and availability assessment of large-scale complex networks. For over 40 years many efforts have been made in developing efficient MC simulation methods and software for determining the confidence interval on system reliability and availability. Classical statistical methods can easily find point estimates of complex system reliability, but it's very difficult and, sometimes, even impossible to obtain confidence intervals of systems reliability and availability.

Simulation can be defined as a technique of performing random experiments on the model that represents a system; it includes sampling stochastic simulation from probability distributions, and involves certain types of mathematical and logical models that describe the behaviour of business or economic system. Simulation is often viewed as to be employed when everything else has failed. Recent advances in simulation methodologies, availability of software, and technical developments have made simulation one of the most widely used and accepted tools in system analysis and operation research.

The Monte Carlo method was used to resolve the Boltzmann equation in the beginning of the *XX century*. In 1908 the famous statistician student *(Gosset)* used the Monte Carlo method for estimating the correlation coefficient in t-distribution.

The term *"Monte Carlo"* was introduced by *John von Neumann* and *Stanislaw Ulam* during World War II as referred by Eckhardt (1987); it was suggested by the gambling casinos at the city of Monte Carlo in Monaco. The Monte Carlo method was then applied to research properties of neutron travel through radiation shielding. The work involved direct simulation of behaviour concerned with random neutron diffusion in fissionable material. After that, Monte Carlo methods were used to complex multidimensional and infinitive integrals and to solve integral equations that were not amenable to analytic solution.

The Monte Carlo method can be used also for solution of deterministic problems. A deterministic problem can be solved by the Monte Carlo method if it has the same formal expression as a stochastic process. There are some different techniques to develop the simulation by Monte Carlo. The next section explains the Hit or Miss Monte Carlo method because is the easiest method to understand (particularly if explained in the kind of graphical language involving a curve in a rectangle).

3.2.1 The Hit or Miss Monte Carlo method

To calculate the one dimensional integral, we must start to assume that the function g(x) is bounded

$$0 \leq g(x) \leq c \ , \ a \leq x \leq b. \tag{3.1}$$

Let Ω denote the area of rectangle in figure 3.2

$$\Omega = \{(x, y) : a \leq x \leq b \ , \ 0 \leq y \leq c\} \tag{3.2}$$

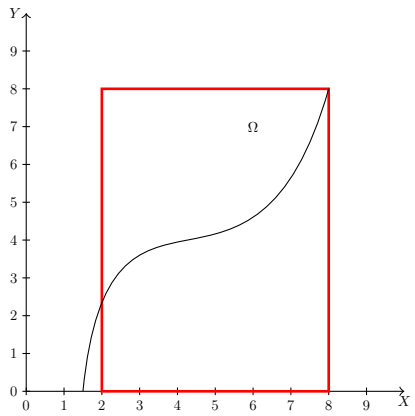


Figure 3.2: Monte Carlo Method - function g(x) and area of Ω

Let (X,Y) be a pair that represents a random vector uniformly distributed over the rectangle Ω with probability density function

$$f_{XY}(x, y) = \begin{cases} \frac{1}{c(b-a)}, & \text{if } (x, y) \in \Omega \\ 0, & \text{Otherwise.} \end{cases} \tag{3.3}$$

The probability p that the random vector *(X, Y)* falls within the area under the curve *g(x)*, denoting $S = \{(x, y) : y \leq g(x)\}$ and observing that the area under the curve *g(x)* is

$$\textit{Area under } g(x) = \textit{area } S = I = \int_a^b g(x)dx \tag{3.4}$$

It's result

$$p = \frac{areaS}{area\,\Omega} = \frac{\int_a^b g(x)dx}{c(b-a)} = \frac{I}{c(b-a)} \tag{3.5}$$

Assume that N independent random vectors $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are generated. The parameter p can be estimated by

$$\widehat{P} = \frac{N_H}{N} \tag{3.6}$$

Where N_H is the number of occasions on which $g(X_i) > Y_i$, i=1,2,...,N, that is, the number of "hits", and $N - N_H$ is the number of "misses"; as depicted in figure 3.2, score a miss if $g(X_i) \leq Y_i$, i=1,2,...,N,

It follows from (3.5) and (3.6) that the integral I can be estimated by

$$I \approx p = c(b-a)\frac{N_H}{N} \tag{3.7}$$

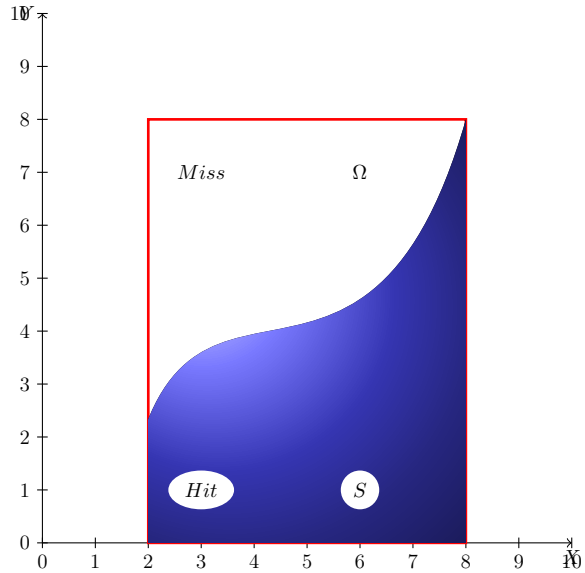


Figure 3.3: Monte Carlo Method - Hit and Miss area

In resume, to estimate the integral I we take a sample N from the distribution (3.3), count the number N_H of hits (below the curve g(x)) and apply (3.7). This approach to approximating integrals is called the Monte Carlo approach.

The algorithm to Hit or miss Monte Carlo method is:

Step 1: Generate a sequence $\{U_j\}_{j=1}^{2N}$ of 2N random numbers.

Step 2: Arrange the random numbers into N pairs $(U_1, U_1^{'}), \ldots, (U_n, U_n^{'})$ in any form such that each random numbers U_1 is used exactly once.

Step 3: Compute $X_i = a + U_i(b-a)$ and $g(X_i)$, i=1,2,...,N.

Step 4: Count the number of cases N_H for which $g(X_i) > cU_i^{'}$.

Step 5: Estimate the integral I and normalized by

$$p = c(b-a)\frac{N_H}{N} \tag{3.8}$$

The appendix A have as subject Monte Carlo methods simulation on reliability and three method are presented and simulated: *Kamat-Riley*, *Rice and Moore, KLMC*. To calculate availability of systems with independent components using Monte Carlo a proposal by Dubi (2000) is introduced.

3.3 Random number generators

When a computer is being used to simulate the real world, random numbers are required to make things realistic. Sometimes it's impractical to examine all possible cases, but with simulation and random sample number it's possible to provide insight into what constitutes the model behaviour. Randomness is also an essential part of optimal strategies in the theory of games and emulate in some way, rolling dice, shuffling decks of cards, spinning roulette wheels, etc., The heart of a simulation is the ability to generate random numbers, where a random number represents the value of a random variable uniformly distributed on (0,1).

The natural chaos that computer graphics schemes often seek to emulate, is itself fake randomness; randomness is after all just a function of available information and not some property of the universe.

Random number generators can be classified into three groups, according to the source of their "randomness": - *TRNGs* true random number generators; - *QRNGs* quasi-random number generators; - *PRNGs* pseudo-random number generators. The true random number generators *(TRNGs)* are hardware solutions such as the method that amplify noise generated by a resistor or a semi-conductor diode and feed this to a comparator. True random number generators can be used for research, cryptography, modelling, etc. The quasi-random number generators *(QRNGs)* is defined as sequentially occupied an solution space. These generators avoids clusters and attempt to evenly fill an *n-dimensional* space with points, without grouping of points in order to obtain as uniform as possible coverage of the domain give up serial independence of subsequently generated values. The pseudo-random number generators *(PRNGs)* are the modern approach using computer to successively generate pseudo-random numbers. These pseudo-random numbers is a sequence of values having the appearances of being independent uniform *(0,1)*. In the next section a more deep study will be describe about generation of pseudo-random number accordingly to some of the classic authors, Rubinstein and Kroese (2016), Ross (2012), Vose (2008) and Gurubilli (2010).

3.3.1 Pseudo-random number generation

The generation of pseudo-random numbers starts with the algorithm with initial value x_0, called the seed, and then recursively computes successive values $X_n, \quad n \geq 1$ by letting

$$x_n = ax_{n-1} \quad modulo\ m \tag{3.9}$$

where a and m are given positive integers, and where the expression modulo m means that ax_{n-1} is divide by m and the remainder is taken as the value of x_n. Each X_n is either $0, 1, ..., m-1$ and the quantity x_n/m- called a pseudo-random number - is taken as an approximation to the value of a uniform (0,1) random variable.

The use of equation (3.9) to generate random numbers is called the multiplicative congruential method. Each of the numbers is called the multiplicative congruential number, it follows that after some finite number a value must repeat itself; and once this happen the whole sequence will begin to repeat. To minimize the risk of repetition, its need to choose the constants a and m, that, for any initial seed x_0, the number of variables that can be generated before this repetition occurs is large.

A guideline that appears to be of help in satisfying the above three conditions is that m should be chosen to be a large prime number, that can be fitted to the computer world size. For 32-bit word machine (where the first bit is a sign bit) it has been shown that the choices of $m = 2^{31} - 1$ and $a = 7^5 = 16807$ result in desirable properties.

Another generator of pseudo-random numbers uses recursions of the type

$$x_n = (ax_{n-1} + c) \ \textit{modulo} \ m \tag{3.10}$$

Such generators are called mixed congruential generators (as they involve both an additive and a multiplicative term). When using generators of this type, often choose m to equal the computer's word length, since this makes the computation of the *modulo* m that is, the division of $ax_{n-1} + c$ by m is quite efficient. Most computer languages already have a built-in random number generator which can be called to generate random numbers.

The first objection to this method, common to all random number generators, is that the succession of X_n values is not at all random. However, we will see later that if we choose the convenient initial parameters, the sequence $\{u_n\}$ can resemble a succession of random numbers.

The second objection is that the values u_i can only take the values $0, \frac{1}{m}, \frac{2}{m}, ..., \frac{m-1}{m}$, so there is no possibility of generating a value between $\frac{2}{m}$ and $\frac{3}{m}$. Choosing large enough $m \geq 10^9$, the set of possible values is sufficiently dense in [0,1] so that the sequence resembles that of a uniform continuous variable in the chosen range.

Example 3.3.1:
Let take the congruential generator $x_n = 5x_{n-1} + 1 mod 9$ with $x_0 = 1$

n		
0	1	–
1	6	0.666
2	4	0.444
3	3	0.333
4	7	0.777
5	0	0

And the solution:
$x_1 = (5 * 1 + 1) mod 9 = 6 mod 9 = 60 \div 9 =$6 with remainder 6

$x_2 = (5 * 6 + 1) mod 9 = 31 mod 9 = 31 \div 9 = 3$ with remainder 4

$x_3 = (5 * 4 + 1) mod 9 = 21 mod 9 = 21 \div 9 = 2$ with remainder 3

$x_4 = (5 * 3 + 1) mod 9 =$ 16mod9 $= 16 \div 9 =$ 1 with remainder 7

$x_5 = (5 * 7 + 1) mod 9 =$ 36mod9$=$ $36 \div 9 =$ 4 with remainder 0

$x_6 = (5 * 0 + 1) mod 9 =$1mod9$=$ $1 \div 9 =$ 1 with remainder 1

The cycle repeat at $x_6 =$ period of 6

In the computer simulation of systems it is assumed it generates a sequence of pseudo-random numbers that can be taken as an approximation to the values of a sequence of independent uniform (0,1) random variables.

3.3.1.1 The middle-square-method algorithm

The middle-square-method was one of the first methods that *John Von Neumann (1946)* used to generate pseudo random numbers. The method is made in order to generate the next number with squaring the current number and let the next number be the middle part of the product.

Step 1: Set a seed number; random-number functions in computers often use dates and times for this purpose.

Step 2: Square the seed.

Step 3: A 8 digit number (the length is *2n*) is necessary, so if the result of square is less than 8 digits, add zeros to the front of the number.

Step 4: Finally, fix the middle four digits; take away the first 2 and last 2 digits and retain the remaining 4 digits and repeat the cycle with this number.

Example 3.3.2:

$$x_0 = 4608 \Rightarrow x_0^2 = 21|2336|64 \Rightarrow x_1 = 2336 \Rightarrow u_1 = 0.2336$$

$$x_1 = 2336 \Rightarrow x_1^2 = 5|4568|96 \Rightarrow x_2 = 4568 \Rightarrow u_2 = 0.4568$$

$$x_2 = 4568 \Rightarrow x_2^2 = 20|8666|24 \Rightarrow x_3 = 8666 \Rightarrow u_3 = 0.8666$$

$$x_3 = 8666 \Rightarrow x_3^2 = 75|0995|56 \Rightarrow x_4 = 995 \Rightarrow u_4 = 0.995$$

$$x_4 = 995 \Rightarrow x_4^2 = 0|995|56 \Rightarrow x_5 = 0995 \Rightarrow u_5 = 0.0995$$

$$x_5 = 995 \Rightarrow x_5^2 = 95|0100|00 \Rightarrow x_6 = 9900 \Rightarrow u_6 = 0.9900$$

$$x_6 = 9900 \Rightarrow x_6^2 = |100|00 \Rightarrow x_7 = 100 \Rightarrow u_7 = 0.100$$

Sometimes the generator ends up in a fairly short cycle. Convergence is a fact of importance in randomness. Once zero is obtained all the following numbers are also zero. Numbers produced pseudo-randomly over some range of values eventually settle down toward a constant number, which is where the random sequence stops. Using a larger n contribute to get a large number of pseudo random numbers before the generator degenerates.

3.3.1.2 Generating continuous random variables

The **inverse transform technique** can't be used to sample all distributions but from some of the most important like the exponential, uniform, Weibull and triangular distribution and the underlying principle for sampling from a wide variety of discrete distributions. Is not the most efficient technique computationally, but the most straightforward.

Consider a continuous random variable having distribution function F. The inverse transformation method for generating a random variable is based on the following proposition:

Proposition 1 (Ross (2012)). *Let U be a uniform (0,1) random variable. For any continuous distribution function F the random variable X defined by:*

$$X = F^{-1}(U)$$

has distribution F.

$F^{-1}(u)$ is defined to be that value of x such that $F(x) = u$.

Proof. Let F_x denote the distribution function of $X = F^{-1}(U)$. Then

$$\begin{aligned} F_X(x) &= P\{X \leq x\} \\ &= P\{F^{-1}(U) \leq x\} \end{aligned} \tag{3.11}$$

F(x) is a distribution function with a monotone increasing function of x and the inequality "$a \leq b$" is equivalent to the inequality "$F(a) \leq F(c)$". Hence, from equation (3.11) yields,

$$\begin{aligned} F_X(x) &= P\{F(F^{-1}(U) \leq F(x))\} \\ &= P\{U \leq F(x)\} \quad \text{since } F(F^{-1}(U)) = U && (3.12) \\ &= F(x) \qquad\qquad \text{since U is uniform(0,1)} && (3.13) \end{aligned}$$

□

The proposition 1 shows that from the continuous distribution function F can generate a random variable X by generating a random number U and then setting $X = F^{-1}(U)$.

Example 3.3.3:
Generate a random variable X having distribution function:

$$F(x) = x^n, \quad 0 < x < 1$$

And let $x = F^{-1}(u)$, then

$$u = F(x) = X^n \quad \text{or, equivalently,} \quad x = u^{1/n}$$

Hence, the procedure can generate such a random variable X by generating a random number U and then setting $X = U^{1/n}$.

One of the most powerful use of the *inverse transform method* is the approach to generating exponential random variables. If X is an exponential random variable with rate *1*, then it's distribution function is given by

$$F(x) = 1 - e^{-x}$$

Let $x = F^{-1}(u)$, then

$$u = F(x) = 1 - e^{-x}$$

or

$$1 - u = e^{-x}$$

or, taking logarithms,

$$x = -log(1 - u)$$

Finally, generate an exponential with parameter *1* by generating a random number U and then setting:

$$X = F^{-1}(U) = -log(1 - U)$$

Noting that $1 - U$ is also uniform on $(0, 1)$ and thus $-\log(1 - U)$ has the same distribution as $-\log U$, that is, the negative logarithm of a random number is exponential distributed with rate *1* and small savings time can be obtained. In addition, if X is exponential with mean *1* then, for any positive c, cX is exponential with mean c. Hence, an exponential random variable X with rate λ (mean $1/\lambda$) can be generated by generating a random number U and setting:

$$X = -\frac{1}{\lambda} \log U$$

3.3.1.3 Continuous distributions without a closed-form inverse

The inverse transform technique for random-variate generation it not possible for a number of useful continuous distributions include the normal, gamma and log-normal distributions because they do not have a closed form expression for their CDF or its inverse. Notice that even a closed-form inverse requires approximation in order to evaluate it on a computer. For example, generating exponential distributed random variates by the inverse CDF $X = F^{-1}(U) = -\ln(1 - U)/\lambda$ requires a numerical approximation for the logarithm function. Thus, there is no essential difference between using an approximate inverse CDF and approximately evaluating a closed-form inverse. The problem with using an approximate inverse CDF is that some of them are computationally slow to evaluate.

To illustrate the idea, suppose a generating a gamma (n, λ) random variable. Since the distribution function F of such a random variable is given by

$$F(x) = \int_0^x \frac{\lambda \exp^{-\lambda y} (\lambda y)^{n-1}}{(n-1)!} dy$$

It is not possible to give a closed form expression for its inverse. However, by using the result that a gamma (n, λ) random variable X can be regarded as being the sum of n independent exponentials, each with rate λ (see section 2.2.3 of chapter). Specifically, generate numbers $U_1, ..., U_n$ and then setting:

$$X = -\frac{1}{\lambda}\log U_1 - \cdots - \frac{1}{\lambda}\log U_n \tag{3.14}$$

$$= -\frac{1}{\lambda}\log(U_1 \ldots U_n) \tag{3.15}$$

where the use of the identity $\sum_{i=1}^{n}\log x_i = \log(x_1, \ldots, x_n)$ is computation time saving and it requires only one logarithm rather than n logarithms computations.

The last result can be used to provide an efficient way of generating a set of exponential random variable by first generating their sum and, then, conditional on the value of that sum, generating the individual values.

For example, to generate X and Y, a pair of independent and identically distributed exponentials having mean *1*, by first generating $X+Y$ and then using the result that, given $X+Y=t$, the conditional distribution of X is uniform on *(0,t)*. The following algorithm can thus be used to generate a pair of exponentials with mean *1*.

Step 1: Generate random numbers U_1 and U_2.

Step 2: Set $t = -\log(U_1 U_2)$.

Step 3: Generate a random number U_3.

Step 4: $X = t\dot{U}_3, \quad Y = t - X$.

Comparing the last algorithm with the direct approach of generating two random numbers U_1 and U_2 and then calculate

$$X = -\log U_1, \quad Y = -\log U_2$$

shows that the above algorithm saves a logarithmic computation at the cost of two multiplications and the generation of a random number.

3.3.1.4 The rejection method

The rejection method for generating a random variable having density function $g(x)$ as the basis for generating from the continuous distribution having density function $f(x)$ by generating Y from g and then accepting this generated value with a probability proportional to $f(Y)/g(Y)$.

Let c be a constant such that:

$$\frac{f(y)}{g(y)} \leq c \quad \text{for all y}$$

Accept the value if $U \leq c$, otherwise reject.

The following algorithm for generating a random variable for the rejection method is:

Step 1: Generate Y having density g

Step 2: Generate a random number U

Step 3: If $U \leq \frac{f(y)}{g(y)}$, set X=Y. Otherwise, return to step 1.

Example 3.3.4 (Ross (2012))**:**
Let use the rejection method to generate a random variable having density function $f(x)$:

$$f(x) = 20x(1-x)^3, \quad 0 < x < 1$$

since this random variable (which is beta with parameter *2.4*) is concentrated in the interval *(0,1)*, let us consider the rejection method with:

$$g(x) = 1, \quad 0 < x < 1$$

To determine the constant c such that $F(x)/g(x) \leq c$, to determine the maximum value of:

$$\frac{f(x)}{g(x)} = 20x(1-x)^3$$

Differentiation of this quantity yields:

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = 20[(1-x)^3 - 3x(1-x)^2]$$

Setting this equal to $\boldsymbol{0}$ shows that the maximum value is attained when $x = 1/4$ and thus:

$$\frac{f(x)}{g(x)} \leq 20\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^3 = \frac{135}{64} \equiv c$$

Hence,

$$\frac{f(x)}{g(x)} = \frac{256}{27}x(1-x)^3$$

and thus the rejection procedure is as follows:

Step 1: Generate random numbers U_1 and U_2

Step 2: If $U_2 \leq \frac{256}{27}U_1(1-U_1)^3$ stop and set $X = U_1$. Otherwise, return to step1

The average number of times that *Step 1* will be performed is $c = \frac{135}{64} \approx 2.11$

3.4 Test for random numbers

The design of *RNG* should involve a rigorous mathematical analysis of their successive values that they produce over their entire period length instead of choose some arbitrary algorithms. After having selected and implemented one *RNG*, must be tested empirically. Statistical tests are also required for *RNGs* based on physical devices like *TRNGs*. The desirable properties of random numbers – uniformity and independence – have to be tested. A huge number of tests can be performed, see L'Ecuyer (2012) and Knuth (2011).

In the beginning the tests applied to *RNGs* were basic and then were followed by more powerful ones to detect regularities in linear generators, for example the tests later proposed by Marsaglia and Tsay (1985). Some of these tests and new ones have been studied more extensively by L'Ecuyer and Simard (2007), Marsaglia et al. (2002) and others.

The number of random tests that can be defined is high and no one can guarantee that a particular generator is fully reliable. Statistical tests can never prove that a *RNG* is infallible and these several tests detect different problems with the *RNGs*. Random tests can contribute to trust or not in the *RNG*.

The good and bad *RNGs* is when the bad *RNG* fail very simple tests, while the good *RNG* fails only very complicated tests.

A brief description of different types of tests discussed in this chapter is as follows:

1. Frequency test. Uses the Kolmogorov-Smirnov or the chi-square test to compare distribution of the set of numbers generated to a uniform distribution

2. Runs test. Tests the run up and down or the runs above and below the mean by comparing the actual values to expected values. The statistic for comparison is the chi-square.

3. Autocorrelation test. Test the correlation between numbers and compares the sample correlation to the expected correlation of zero.

4. Counts the number of digits that appear between repetitions of a particular digit and then uses the Kolmogorov-Smirnov test to compare with the expected size of gaps.

There is RNGs for all types of applications and they are designed so that, their output sequence is a truly image of a sequence of independent uniform random variables, usually over the real interval (0, 1).

In testing for uniformity, the hypothesis are as follows:

$$H_0: \quad R_i \sim U[0,1]$$

$$H_1: \quad R_i \nsim U[0,1]$$

The null hypothesis, H_0 is verified if the numbers are distributed uniformly on the interval $[0, 1]$. Not reject the null hypothesis means that no evidence of non-uniformity has been detected on the basis of this test. This does not imply that further testing of the generator for uniformity is unnecessary.

Testing for independence, the hypothesis are as follows:

$$H_0: \quad R_i \sim independently$$

$$H_1: \quad R_i \nsim independently$$

This null hypothesis, H_0, verified if the numbers are independent. Not reject the null hypothesis means that no evidence of dependence has been detected on the basis of this test.

For each test, a level of significance α must be stated. The level α is the probability of rejecting the null hypothesis given that the null hypothesis is true, or

$$\alpha = P(\text{reject } H_0 | H_0 \text{ true})$$

The decision sets the value of α for any test. Frequently, α is set to *0.01* or *0.05*.

If several tests are conducted on the same set of numbers, the probability of rejecting the null hypothesis on at least one test [i.e., making a typeI (α) error], increases. If the level of significance is $\alpha = 0.05$ and five different tests are conducted on a sequence of numbers; the probability of rejecting the null hypotheses on at least one test, by chance alone, may be as large as 0.25%, L'Ecuyer and Simard (2007).

When applying a test of hypothesis, in classical statistics, it's necessary to select beforehand a rejection area R whose probability under H_0 equals the target test level (e.g., *0.05 or 0.01*), and reject H_0 if and only if $Y \in R$. This procedure it's the most appropriate when have a fixed and small sample size, but is not the best approach in the field of RNG. Indeed, the sample sizes of RNG are huge and can usually be increased at will. To validate the test simply compute and report the *p-value* instead of selecting a test level and a rejection area.

$$p = P[Y \geq y | H_0]$$

where y is the value taken by the test statistic Y. If Y has a continuous distribution, then p is a *U(0, 1)* random variable under H_0.

For randomness tests, the value of p if it is close to *1* can be seen as a measure of uniformity and the generator is producing the values with excessive uniformity, and if the value of p is close to *0*, it can be understood also as a measure of uniformity.

The RNG fails the test, if the value of p is very close to *0* or *1*, but this value is suspect and does not clearly indicate the rejection ($p = 0.003$, for example), so it is better to

repeat. The test can be replicated the number of times that is necessary. This approach is possible because there is usually no limit (beyond computational processing time) on the amount of data that can be produced by an algorithm to ensure that the developed RNG is acceptable or not.

When a generator begins to fail decisively in a test, the *p-value* of the test usually converges to *0* or *1* in an exponential way depending on the size of the test sample. One way to address this trend will be to increase the sample size or review the generator algorithm.

3.4.1 The Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test measure the degree of agreement between the distribution of a sample of generated random numbers and the theoretical uniform distribution. The test is based on the null hypothesis of no significant difference between the sample distribution and the theoretical distribution.

This test compares the continuous *CDF* of the *Uniform* distribution to the empirical *CDF*, $U_n(x)$, of the sample of N observations.

If $U_1 \leq U_2 \leq \cdots \leq U_N$ are the N observations sorted by increasing order. As N becomes larger, $U_n(x)$ should become a better approximation to $F(x)$, provided that the null hypothesis is true.

The Kolmogorov-Smirnov test is based on the largest absolute deviation between $F(x)$ and $U_n(x)$ over the range of the random variable. That is based on the statistic:

$$D = max|F(x) - U_n(x)|$$

The sampling distribution of D is known and is tabulated as a function of N in table.

For testing against a uniform *CDF*, the test procedure follows these steps:

Step 1: Rank the data from smallest to largest, so that:

$$U_1 \leq U_2 \leq \cdots \leq U_N$$

Step 2: Compute

$$D^+ = \underset{1 \leq i \leq N}{max} \left\{ \frac{i}{N} - U_i \right\}$$

$$D^- = \underset{1 \leq i \leq N}{max} \left\{ U_i - \frac{i-1}{N} \right\}$$

Step 3: Compute $D = max(D^+, D^-)$.

Step 4: Determine the critical value, D_α, from table for the specified significance level α and the given sample size N.

Step 5: If the sample statistic D is greater than the critical value D_α, the null hypothesis that the data are a sample from uniform distribution is rejected. If $D \leq D_\alpha$, conclude that no difference have been detected between the true distribution of $U_1, U_2 \cdots U_N$ and the uniform distribution.

Example 3.4.1:
Suppose that the five numbers $0.76, 0.6, 0.9, 0.04, 0.2$ were generated and it is desired to perform a test for uniformity using the KS - test with a level of significance α of **0.05**.

First, the numbers must be ranked from the smallest to largest. The calculations can be facilitated by use the table 3.1; the top row lists the numbers from smallest U_1 to largest U_5.

The computation for D^+, namely $i/N - U_1$ and for D^-, namely $U_i - (1-i)/N$ are easily accomplished using table 3.1.

The statistics are computed as $D^+ = 0.2$ and $D^- = 0.2$. Therefore,

$$D = max(0.2, 0.2) = 0.2$$

The critical value of D_α, obtained from tabulated values for $\alpha = 0.05$ and $N = 5$, is *0.565*. Since the computed value, $D = 0.2$, is less than the tabulated critical value, *0.565*, the hypothesis of no difference between the distribution of the generated numbers and the uniform distribution is not rejected.

U_i	0.04	0.2	0.6	0.76	0.9
i/N	0.2	0.4	0.6	0.8	1.00
$i/N - U_i$	0.16	0.20	0	0.04	0.1
$U_i - (i-1)/N$	0.04	0	0.2	0.16	0.1

Table 3.1: Kolmogorov-Smirnov Test - Matrix Example

3.4.2 Runs test - up and down

The runs test examines the arrangement of the numbers in a sequence to test the hypothesis of independence.

A run is defined as succession of similar numbers preceded and followed by a different number. The length of the run is the number of events that occur in the sample.

There are two parameters in a runs test. The first one is the number of runs and the length of runs is the second parameter. The types of runs counted in the first case might be runs up and runs downs. An up run is a sequence of numbers each of which is succeeded by a larger number. Similarly, a down run is a sequence of numbers each of which is succeeded by a smaller number.

If a is the total number of runs in a truly random sequence, the mean and the variance of a is given by

$$\mu_a = \frac{2N-1}{3} \tag{3.16}$$

and

$$\sigma_a^2 = \frac{16N-29}{90} \tag{3.17}$$

For $N > 20$, the distribution of a is reasonably approximated by a normal distribution, $N(\mu_a, \sigma_a^2)$. This approximation can be used to test the independence of numbers from a generator. In that case the standardized normal test statistic is developed by subtracting the mean from the observed number of runs, a, and dividing by the standard deviation. That is, the test statistic is

$$Z_0 = \frac{a - \mu_a}{\sigma_a}$$

Substituting equation 3.16 for μ_a and the square root of equation 3.17 for σ_a yields

$$Z_0 = \frac{a - [(2N-1)/3]}{\sqrt{(16N-29)/90}}$$

where $Z_0 \sim N(0,1)$. Failure to reject the hypothesis of independence occurs when $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$, where α is the level of significance.

3.4.3 Wald-Wolfowitz test

The Wald-Wolfowitz test, also called as the Runs test for randomness, is used to test the hypothesis that a series of numbers is random. A run is a set of sequential values that are either all above or below the mean. The data are first centred about their mean and then the total number of runs is computed along with the number of positive and negative values. A positive run is then a sequence of values greater than zero, and a negative run is a sequence of values less than zero. After that, the test is performed to see if the number of positive and negative runs are distributed equally in time.

Let n and m be the number of individual observations above and below the mean and let b be the total number of runs. The maximum number of runs is $N = n + m$, and the minimum number of runs is one. Given n and m, the mean and the variance of b for a truly independent sequence are given by:

$$\mu_b = \frac{2nm}{N} + \frac{1}{2} \tag{3.18}$$

and

$$\sigma_b^2 = \frac{2nm(2nm-N)}{N^2(N-1)} \tag{3.19}$$

For either n or m greater than *20,* b is approximately normally distributed. The test statistic can be formed by subtracting the mean from the number of runs and dividing by the standard deviation, or

$$Z_0 = \frac{b - (2nm/M) - 1/2}{\left[\frac{2nm(2nm-N)}{N^2(N-1)}\right]^{1/2}}$$

Failure to reject the hypothesis of independence occurs when $-Z_{\alpha/2} \leq Z_0 \leq Z_{\alpha/2}$, where α is the level of significance.

3.4.4 Mann-Kendall Test

The non-parametric Mann-Kendall test is used to detect monotonic trends in series of data, like enviroment, biological, reliability, etc... The null hypothesis, H_0, is that the data come from a population with independent realizations and are identically distributed. The alternative hypothesis, H_1, is that the data follow a monotonic trend. The Mann-Kendall test statistic is calculated according to:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn(X_j - X_k)$$

With

$$sgnx = \begin{cases} 1 & if \quad x > 0 \\ 0 & if \quad x = 0 \\ -1 & if \quad x < 0 \end{cases}$$

The mean of S is $E[S] = 0$ and the variance σ^2 is

$$\sigma^2 = \left\{ n(n-1)(2n+5) - \sum_{j=1}^{P} t_j(t_j - 1)(2t_j + 5) \right\} /18$$

where p is the number of the tied groups in the data set and t_j is the number of data points in the jth tied group. The statistic S is approximately normal distributed provided that the following Z-transformation is employed:

$$Z = \begin{cases} \frac{S-1}{\sigma} & if \quad S > 0 \\ 0 & if \quad S = 0 \\ \frac{S+1}{\sigma} & if \quad S < 0 \end{cases}$$

The statistic S is closely related to Kendall's τ as given by:

$$\tau = \frac{S}{D}$$

where

$$D = \left[\frac{1}{2}n(n-1) - \frac{1}{2}\sum_{j=1}^{P} t_j(t_j-1)\right]^{1/2} \left[\frac{1}{2}n(n-1)\right]^{1/2}$$

The Mann-Kendall trend test has very interesting features. The Mann-Kendall trend test has 0.98 efficiency relative to the usual least squares method of testing $\beta = 0$. An empirical simulation study of Hipel et al. (1986) showed that the Mann-Kendall test have a good performance for detecting a variety of deterministic trends such as a step-intervention or a linear trend.

In the case of no ties in the values of $Zt, t = 1, \cdots, n$ the Mann-Kendall rank correlation coefficient τ has an interesting interpretation. In this case, the Mann-Kendall rank correlation for a trend test can be written

$$\tau = \frac{S}{\binom{n}{2}}$$

where

$$S = 2P - \binom{n}{2}$$

where P is the number of times that $Zt_2 > Zt_1$ for all $t_1, t_2, \cdots, t_n$ such that $t_2 > t_1$. Thus $\tau = 2\pi c1$, where πc is the relative frequency of positive concordance, i.e., the proportion of time for which $Zt_2 > Zt_1$ when $t_2 > t_1$. The relative frequency of positive concordance is given by $\pi c = 0.5(\tau + 1)$. The Mann-Kendall test is essentially limited to testing the null hypothesis that the data are independent and identically distributed.

3.4.5 Turning Point Test

In statistical hypothesis testing, a turning point test is a statistical test of the independence of a series of random variables. Maurice Kendall and Alan Stuart describe the test as *"reasonable for a test against cyclicity but poor as a test against trend"*. The test was first published by *Irénée-Jules Bienaymé* in 1874. The turning point test for randomness is used to determine if the peaks and troughs (or turning points) of a serial data set (time-series) is independent of the order of the observations. This test requires that the sample size be > 15 and that the measurements were obtained under similar conditions, Marsaglia and Tsay (1985).

Let T be the number of turning points then for large n, T is approximately normally distributed with mean $(2n4)/3$ and variance $(16n29)/90$.

The test statistic,

$$Z = \left| \frac{t_p - \overline{x}}{s} \right|$$

Where,

$$t_p = peaks + troughs$$
$$\overline{x} = \frac{2}{3}(n-2)$$
$$s = \frac{16n - 29}{90}$$

3.4.6 Bartels test

The rank version of Von Neumann's ratio test for randomness is Bartels test.

Bartels test ranks all the samples from the smallest to the largest and compare the magnitude of each observation with its preceding samples.

The corresponding sequential number of Xi: R(Xi) suppose rank arrangement from all n! possibilities should be equip-probable.

The test statistic RVN is

$$RVN = \frac{\sum_{i=1}^{n-1}(R_i - R_{i+1})^2}{\sum_{i=1}^{n}(R_i - (n+1)/2)^2}$$

where, $R_i = rank(X_i); \quad i = 1 \ldots . n.$

It is known that $(RVN - 2)/\sigma$ is asymptotically standard normal, where,

$$\sigma^2 = \frac{4(n-2)(5n^2 - 2n - 9)}{5n(n+1)(n-1)2}$$

Chapter 4

Model data and simulation

In lifetime data analysis on reliability of equipment it's impossible to follow all production units until the end of their lifetimes. Censored data appears when there is some information about the units, but don't know the exact time of failure or death. The experimental observation period is defined as the time elapsed, since the study or experiment begin ($t = 0$) until it is terminated (time T_0). However,sometimes appear the need to discontinue the study before all the equipment in the study experience the failure.

The analyst stops collecting the data, and do the analyses with data that have been observed. In such cases, the experiment has been "suspend", "censored" or "truncated". For example, it is not feasible to follow a set of electric lamps till all of them fail. In this point of the study, the data may be a mixture of two different types of observations. For some items, failure may not have occurred, while for some others it may have. The items which have not failed, when the study is stopped, are said to be censored. Censoring is a property of the sample, and an practical constraint because of the experimenter is unable to know the beginning of their lifetime or to follow the units till the end of their lifetimes.

4.1 Censoring and truncation

Censoring data can arise from many reasons, and depending on the reasons, censoring may be of many different types. For detailed account of censored data, one may refer to Mitra (2012). Some types of censoring are describe:

1. observed, truncated, or censored
2. left, right, or interval censored
3. Type I, Type II, or randomly censored
4. single or multiple censoring values

The data is considered complete when it is known the exact time of each system failure. In many cases the data contain uncertainties, i.e., it is not known the exact moment when

the failure occurred. The data containing such uncertainty when the event occurred are regarded as incomplete or partial. Incomplete data can be classified into censored or truncated as explained by Gijbels (2010).

Censoring, from the theoretical point of view, may not be the most efficient way to conduct an experience, but, due to time, cost or practical aspects, it's so frequent that researchers had to find ways to deal with it.

A characterization of the censoring mechanisms it's important to better analyse the data and the phenomena in study. Such characterization can be based on several elements as the status of the entity observed, the span of the study, the dynamic of the system in study and the start time and the finish time of our observation. Censoring mechanisms can also be characterized based on when and how is defined the time to finish the study.

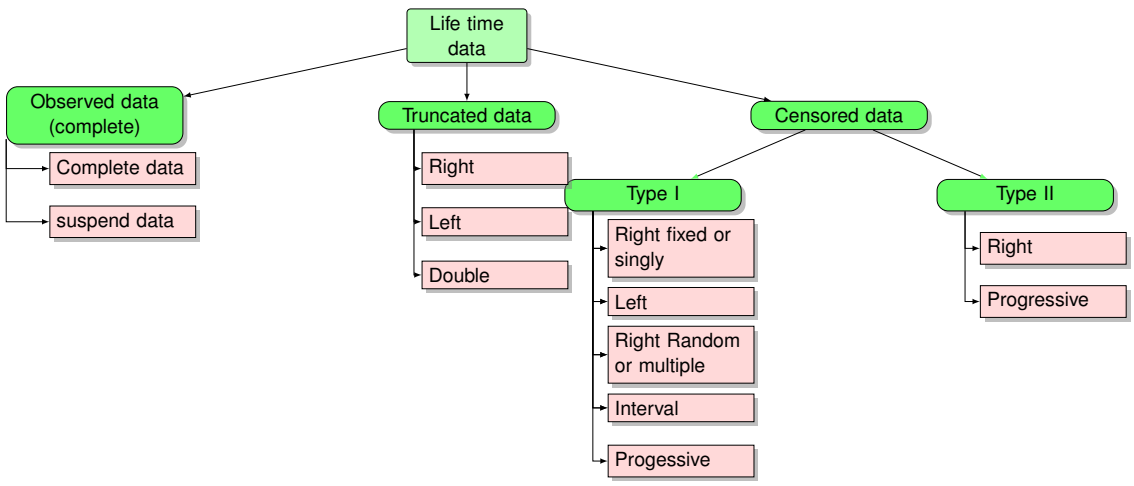


Figure 4.1: The structure and classification of type of data in reliability

Censoring can be informative or non-informative. The analysis methods discussed and assumed in classic research, is that, censoring mechanism must be independent of the survival mechanism or lifetime, because when the cause of censoring is not associated to the lifetimes of the item or biologic unit, the censoring is said to be non-informative. The normal analysis of censored data are based on the assumption that the censoring is non-informative. When the censoring cause is related to the lifetimes, it is said to be informative censoring.

For example, in reliability, when a experience observes that some component or equipment are not operating properly and may fail shortly, then its better remove those components from the experience. In this case, the cause of censoring is directly related to the lifetimes of those items. The experimenter take out those units knowing if they were kept during the test, those units may have failed. This is an example of informative censoring.

In medical study a patient may withdraw from an experience when choose another treatment or stop that treatment for another clinical issues, this is informative censoring. But, if the patient exits from the medical study due to some other reasons, for example,

if moving to another region (i.e., the reason for withdrawal is not related to his lifetime), then the censoring becomes non-informative.

4.1.1 Complete data or observed data

Complete data is defined when the event is observed and known exactly the value of each sample unit that are in the study. For example, in reliability the event is the real time to failure observed for all n units in our sample. If four units are tested, beginning in $t = 0$ and all fail during the test and record all time to failure, in that situation the study have complete information and complete data (see figure 4.2).

Let the n lifetimes of the size-n sample be $T_1, ...T_n$. Their order statistics are:

$$T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(n)}$$

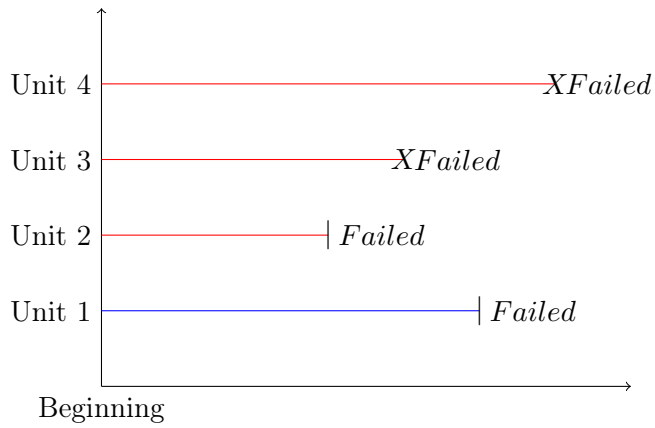


Figure 4.2: Type of data: Complete Data

4.1.2 Type I or time censoring

Censoring schemes are referred to as time censoring, where the end of the observation period is not determined by an event of interest (e.g., failure). Such type of censoring are not event-driven and are known as type-I.

With respect to the status of the entity observed, censoring can occur at either extreme (or at both ends) of the entity life. That is, the researchers don't know exactly at what time the life of the entity started or finished. The reason may be because the entity in question may have already failed at the time the observation begin or the life may have not yet finished (e.g., failed) by the time final observation period.

To better analyse this complex issue some figures will be used to illustrates the censoring situation.

Incomplete data give only part of the information about the failure time of the units under review. However, this information should not be ignored or treated as failure. In

the absence of such data, it would not be possible to make good estimation parameters and, after that, make a proper analysis.

4.1.3 Fixed Right Type I

This type of censoring occurs when a study is designed to end after C years/hours/units of follow-up. In this case, all items who does not have an event observed during the course of the study is censored at C years/hours/units.

One of the most common types of censored data, which may arise in real cases, is fixed right type-I censored data. For fixed right type-I censored data, all units of a system are observed up to the date of completion of the study. For this censorship scheme the time each unit is under observation is fixed, while the number of units that fail (uncensored observations) is random.

In these type of censoring, the experiment stopping time (T_0) is pre-established and the number of failures observed during the period of experimentation is random.

Putting an end to the experiment and stop monitoring all the entities, at some pre-specified time T_0, which is independent of the event of interest.

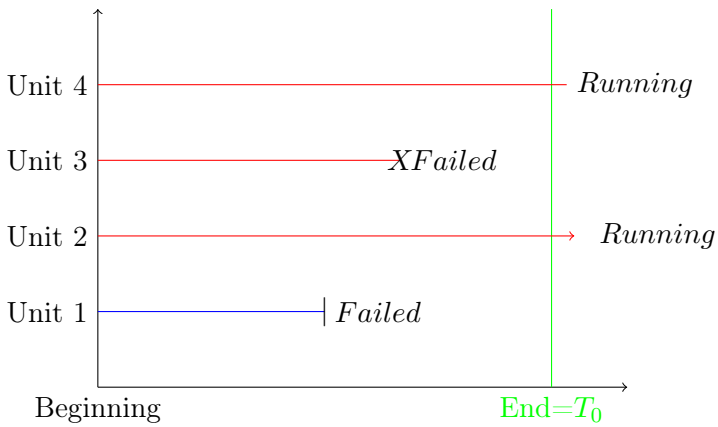


Figure 4.3: Type of data: Fixed Right Type I

4.1.4 Random right Type I

The type of censoring known as right censoring can be observed in figure that shows an entity that has been monitored since the beginning of life (i.e., at the start of the experience) but which have ceased to observe before the experience end (time T_0) or its failure.

Random right type I censoring is designed when the study finish after C years, but censored subjects do not all have the same censoring time, or the entity is observed for some time, after which the research is not able to monitor it any more. The reasons why random right type I censoring might occur:

- a entity does not experience the event before the study ends;

- a person is lost to follow-up during the study period;
- a person withdraws from the study

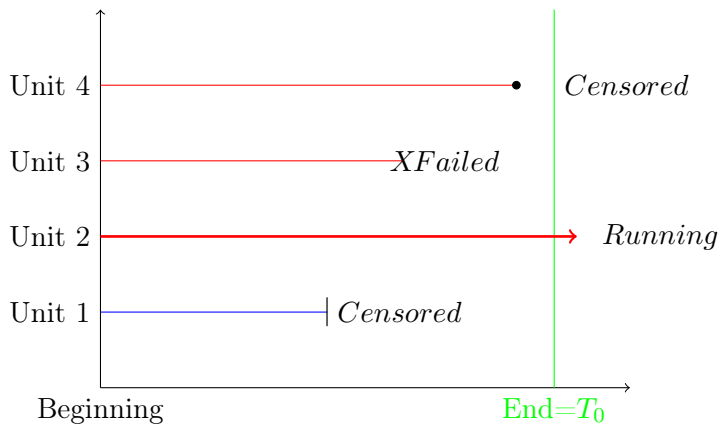


Figure 4.4: Type of data: Random Right Type I

The "x" symbols in figure 4.4 represent the time of start or finish monitoring the censored entities. The study have beginning (of entity life, at time zero) and end in the final of the experimental observation period (time T_0).

4.1.5 Left censored type I

In left censored data, a failure time is only known to be before a certain time. Like right censoring, the censoring in time can come from left also. In this case, it is known that a unit failed before some pre-specified time C or, in general, C_i, but the actual time of failure of the left censored unit is unknown.

For example a certain unit failed before *100* hours but don't know exactly when. In other words, it could have failed any time between *0* and *100* hours. This is identical to interval censored data in which the starting time for the interval is zero.

For left censoring, those are the same concept. The censoring time is the time started to observe. For right censoring, that's reversed, the censoring time is when stop observing.

In figure 4.5 the cases in *unit 2* and *4* that shows an unit that has already been "operating" for some unknown period of time, before start monitoring it.

4.1.6 Interval censored type I

Interval censoring type I happens when only the number of failures in specific intervals are available, without any more specific information about the individual failure times. In figure 4.6 is possible to see some examples of interval censored

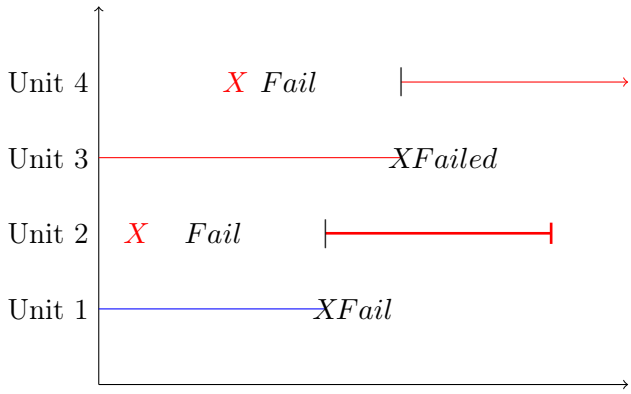


Figure 4.5: Type of data: Left Type I

Interval censored data reflects uncertainty of the exact times the units fails within an interval. This type of data frequently comes from tests or situations where the objects of interest are not constantly monitored.

For example, a test in five units with inspection each 100 hours, the report is only a unit failed or did not fail between inspections. The beginning and the end of the entity life are now unknown, it's only monitor for some intermediate part of its span "life". This type of censored data is also called inspection data by some authors.

It's recommended to avoid interval censored data because they are less informative compared to complete data. However, there are cases when interval data have to be used due to the nature of the product, the test and the test equipment. In those cases, caution must be taken to set the inspection intervals to be short enough to observe the dispersion of the failures. For example, if the inspection interval is too long, all the units in the test may fail within that interval, and thus no failure distribution could be obtained.

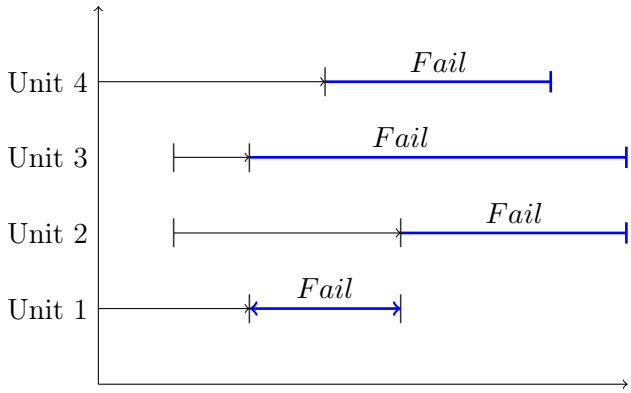


Figure 4.6: Type of data: Interval Type I

4.1.7 Type II or failure censoring

Type-II censoring in statistical literature is the type of censoring that have a number of failures to be observed fixed at start of the experiment. For example, an experience starts

with n units, and the research decides to continue the experience until the rth failure occurs, i.e., until time $T_{r:n}$, which is the rth smallest statistic order among the lifetimes of these n units. When the experiment stops, a pre-specified number of failures r is obtained, and all that is known about the censored observations is that their lifetime $T > T_{r:n}$.

This type of censoring is used in many reliability fields when there is a pre-specified number of events.

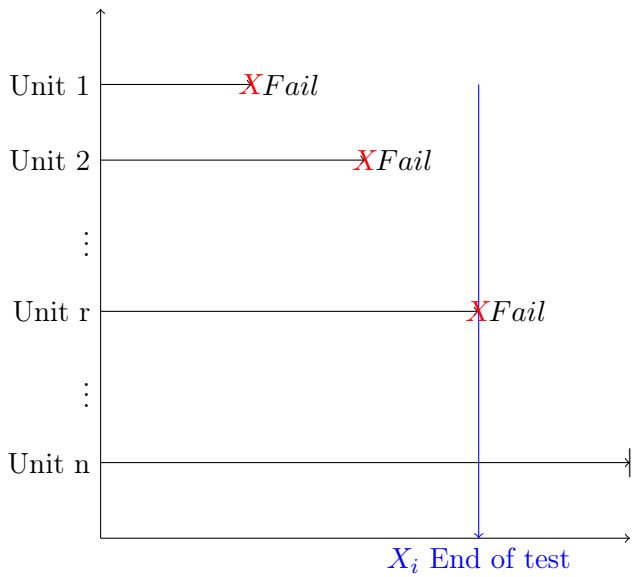


Figure 4.7: Type of data:right type II or failure censoring

In figure 4.7 is observed a sample of n entities until the time of occurrence of some pre-specified event of interest, such as the time of the ith failure or death ($i \leq n$) designated by the X_i in figure 4.7. That is:

$$0 < x_1 < x_2 < ... < x_n < \infty$$

At the time of the ith failure (failure times X_i are denoted in the figure 4.7 by an $\boldsymbol{X}$) the observation of the n_i sample elements remaining in operation. This other censoring scheme is often referred to as "failure" or "event" truncation in these cases, the experiment stopping time (X_i) is random and the number of failures i occurred during experimentation is pre-established.

The number of failures or events of interest observed during the work of the equipment or the experience is less than the total of entities in censoring scheme (type I or II). The statistical distribution of failures of the equipment have to be estimated, and after that, the probability $''p''$ of occurrence of an event during the observation period (time t_0), can be calculated and the model of underlying life distribution and the parameters estimation of interest such as MTTF - Mean Time To Failure , failure rate, etc.

In real life, in the production plant, the problem of modelling "life" is further complicated and need different approach. For example, different situations are when components

are replaced in the same time of occurrence or when the distribution of reliability is not exponential or Poisson homogeneous and, in such cases, the hazard function is time-dependent and there are several additional parameters to estimate from the data. Other situation, having more complex censoring mechanics creates many more theoretical difficulties.

Time censored experiments is a test which terminated at a pre-specified time t_0 as opposed to test that have defined the number of failures. In this test, time censored, the total operating time T of all equipment in operation and the total number of failures K are known but the individual failure times of equipment are not all known . Time censored estimation is approached in different ways, depending on operational or experimental conditions and the nature of the data.

4.2 Simulation data and software

Simulation studies should be designed with rigour, similar to the study of real data, since they should represent the results of real events. Simulating data sets requires an assumed distribution for the data and full specification of the required parameters.

Actually, there is a very large number of tools and simulation software, for a huge variety of applications, from the medical field, to *FEA* - finite elements analysis , hydraulics and robotic simulations, etc.

In reliability there is a lot of simulation studies and specifically in censored data using Weibull distribution. In these studies very few published sufficient details to assess the integrity of the study design or to allow readers to fully understand all the processes required and explain how they have developed the algorithm and the program of simulation and how they generate the random numbers.

Random numbers are the essential basis of the simulation. Usually, all the randomness involved in the model is obtained from a random number generator that produces a succession of values that are supposed to be realizations of a sequence of independent and identically distributed *(i.i.d.)* random variables, L'Ecuyer and Simard (2007).

This section briefly describes the software used in the *PhD*. All of them can be considered a very productive tool in the area of programming and simulation, but the best to work and have been most suitable for the models and simulation of the PhD research was the software R.

4.2.1 Simulation with Program R

The software *R* is a language for statistical computing and graphics and is available as free software under the terms of the Free Software Foundation's the *GNU* General Public License in source code form. Software *R* provides a wide variety of statistical (linear and non-linear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible and is often the vehicle of

choice for research in statistical methodology, and *R* provides an Open Source route to participate in that activity

R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes:

- an effective data handling and storage facility;
- a suite of operators for calculations on arrays, in particular matrices;
- a large, coherent, integrated collection of intermediate tools for data analysis,
- a graphical facilities for data analysis and display, either on-screen or on hard-copy;
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.

The PhD used the Integrated Development Environment - *IDE* to program in *R* called *Rstudio*, which is an open source and includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

4.2.2 Simulation with python and anaconda

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level, built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

RODEO is an open source python *IDE*, is a development environment that is lightweight, intuitive and yet customizable to its very core. It is an *IDE* that has been built especially for data science/Machine Learning in Python and Rodeo and is also known as the *RStudio* clone as uses the Ace Editor as its under lying layer, just the same as what powers *RStudio*.

Rodeo is a native Python IDE that can be of help to its users in quickly get some idea about data structures without having to write any additional lines of code hence reducing the time required and also has some basic package management and plotting views.

4.2.3 Simulation with Matlab

MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment and proprietary programming language developed by MathWorks. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C, Java, Fortran and Python.

Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing abilities. An additional package, Simulink, adds graphical multi-domain simulation and model-based design for dynamic and embedded systems.

MATLAB was first adopted by researchers and practitioners in control engineering, but quickly spread to many other domains. It is now also used in education, in particular the teaching of linear algebra, numerical analysis, and is popular amongst scientists involved in image processing.

4.3 Simulation data censored with statistical distribution

The advances in technology have enable simulation studies to be more accessible. Performing simulations is not simple. Many decisions are required prior to the commencement of simulations, but there is, generally, no one simple correct answer to the problem. A brief search of published papers in reliability censor data area, suggested that the majority of simulation studies reported in the literature are not providing sufficient details of simulation random generator data. Modifications of the simulation process, such as altering the number of simulations or collecting additional parameters or another parameters are possible, but can be time-consuming. Ideally, when testing a model in reliability for simulation, the algorithm to generate random variables of practical interest can have a bad structure interference between the RNG and the simulation model that will show up in the results. Finally, is important to test the simulation algorithms and the influence of the change of parameters.

Several methods and techniques have been proposed for analysing different types of reliability data over the past decades. Most of them, refer to complete data. However, evaluation of highly censored reliability data has not been widely studied. Nelson (1985) presented an excellent work on this topic. In the beginning, few of the studies uses simulation tools, but along the time the use of simulation in reliability field increase, most of them to estimation parameters.

Olteanu and Freeman (2010) conducted a simulation study that compared the performance of maximum likelihood (ML) and median-rank regression (MRR) methods in estimating Weibull parameters for highly censored reliability data.

In addition to the well-known large-sample optimal properties associated with ML estimators, studies shown that ML estimators are generally hard to beat consistently even in small samples, Genschel and Meeker (2010) and Somboonsavatdee et al. (2007).

Recently, the estimation of parameters from different lifetime distributions based on progressive Type-II censored samples are studied by several authors including: - Childs and Balakrishnan (2000), Balakrishnan and Kannan (2001), Mousa and Jaheen (2002) and Soliman (2005).

Many article use $\%C$ - percentage of data censored to compare and analysed the model and the results of study simulations, like in Balakrishnan and Mitra (2012), Birolini (2017) and Ross (2012). The use and application of data censored in the field of reliability can be see in Wang and Coit (2005) and Horst (2009).

Burton et al. (2006) proposed to generate a random non-informative right censoring with specific proportion of censored observations in a similar manner to the uncensored survival times by assuming a exponential distribution for the censoring times, but can be Weibull or uniform without including any covariates. For Burton et al. (2006), it's by iteration that the parameters of censoring distribution will be achieved. Halabi and Singh (2004) in other way, provide formulas for determining parameters for standard survival and censoring distribution. The censoring mechanism can also be extended to incorporate dependent informative censoring.

4.4 An algorithm to simulation right type I censored data

A fundamental part of any simulation is the algorithm used to generate the random numbers. The random number generation is the "Mersenne-Twister", from *Matsumoto and Nishimura (1998)*. A twisted GFSR with period 219937^{-1} and equidistribution in 623 consecutive dimensions (over the whole period). The "seed" is a 624-dimensional set of 32-bit integers plus a current position in set that use the *Mersenne-twister* RNG.

In the simulation studies it's essential to define how the results will be stored after each simulation, in order to avoid the risk to repeat the simulations. The estimate of interest in this study will be the estimate t_c ***- time censoring***. the sample was *50* and *1000*. The routine is made M times (in this case 1000) and then calculate the mean:

$$\mu = \sum_{i=1}^{m} \frac{t_{c_i}}{m}$$

as a measure of the true estimate of interest.

It's fundamental to measure the uncertainty in the estimate of the parameter t_c which represents the percentage of $C\%$ data censored. The empirical standard deviation σ is calculated as the standard deviation of the estimates of interest from all simulations (in

this case M=1000)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{m}(t_{c_i} - \mu)^2}{m}}$$

The average of the estimated within study simulation, σ could be used. Increasing the number of simulations will reduce the SE (σ) of the simulation process, i.e., $\sigma(t_c)/\sqrt{m}$, but this will be computational expensive and therefore variance reduction techniques could be used.

After the simulation have been performed, it's necessary to define the criteria for evaluating the results obtained from the different scenarios or statistical approach. The study use the change of parameters of each distribution.

The simulated result is compared with the true values and provides a measure of the performance and precision of the algorithm in study.

The estimates of simulations is the main reason and hence the average of estimates over all simulation is used to calculate accuracy measures. different scenarios and models was analysed. There is a trade-off between the amount of bias and the dispersion or variability. Some authors argue that having less bias is more crucial than producing a valid estimate of sampling variance. However, models or methods that result in a biased estimated with little variability may be considered not so accuracy or conversely an unbiased estimate with large variability.

To evaluate the performance of statistical methods and algorithms with different distribution parameters it's used the indicators Mean SE(σ) (MSE) and the Percentage Error (PE) - associated with the estimated of each time censoring t_c. The PE use the estimate t_c and the theoretical T_C computed by the following relation:

$$PE_{T_v} = \xi = \frac{\left|t_c - T_{c(exact)}\right|}{T_{c(exact)}} x100$$

The model optimize the simulation and give very good results. It's necessary to calculate the time of censoring from each statistical distribution; generically is to do the inverse function of PDF function to calculate the time censoring t_C and put this value in the algorithm of simulation. This procedure reduces the time consumed in computation and with a large sample is very precise and comes closest to the percentage of censored data defined (theoretic).

The algorithm to generate random censor data have these steps:

Step 1: Define initial parameters and values to simulation

Step 2: Calculate t_c (with parameters of distribution and number $C\%$ of censored data and the number of sample n)

Step 3: Calculate the order of ith number that begin the censor data of a sample (censoring-order i_c).

Step 4: Generate the vector Y that represent t_i random times from distribution model

Step 5: Order the vector Y

Step 6: Find the time for censoring-order - $Y(i_c)$

Step 7: Repeat M times from step 3 to 6 and save to vector T_{cens}

Step 8: From vector T_{cens} calculate the mean and standard deviation

Step 9: From step 8 calculate the error $\xi = |T_c - \mu|$

Step 10: Plot the two function and save the results to table

An example (partial) of program with Weibull distribution is show in algorithm 1. The other distribution is reuse and similar to calculate t_C. The first part of the program in R software is:

```
for (i in 1:length(cen)){
   censoring_order <- n*(1-cen[i])# order in the vector censored
        for (j in 1:length(beta)){
                  #Execute m=100 cycles to obtain
                for (k in 1:m){
                        y <- rweibull(n,beta[j],wscale)
                        T_order <- sort(y) # sort by ascending
                        T_cens[k] <- T_order[censoring_order]
                }
        t <- wscale*(-log(cen[i]))^(1/beta[j]) # time censor
        #obtain the mean and the standar deviation of sample
        ysdt <- sd(T_cens) # standard deviation of sample
        ymeant <- mean(T_cens) # mean of the sample
        #obtain the mean and the standar deviation of sample
        wsdt[j] <- sd(T_cens) # standard deviation of sample
        wmean[j] <- mean(T_cens) # mean of the sample
        werf[j] <- abs(t-wmean[j])/t*100
        printgraf(n,wscale) #function to print the functions
```

Algorithm 1: Program in R from Weibull distribution (partial)

The flow chart that resume the general proposal model can be like the figure 4.8.

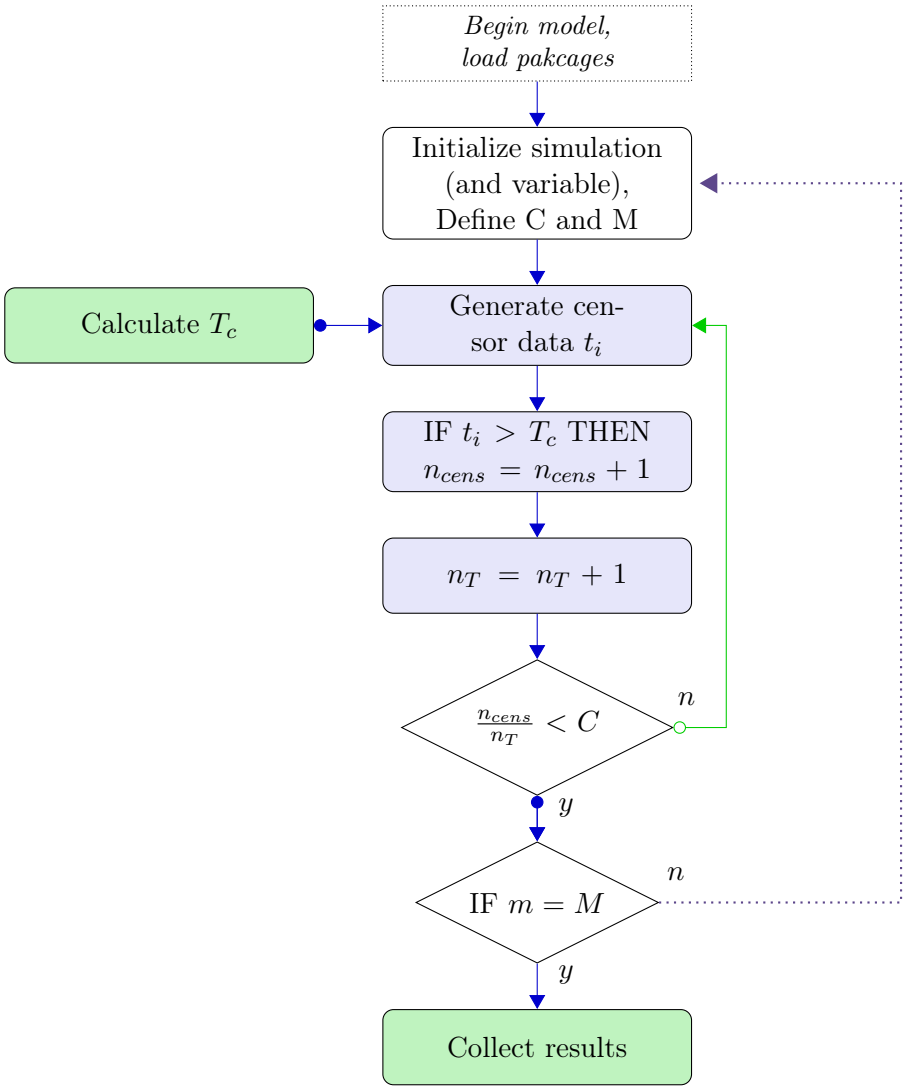


Figure 4.8: Flowchart of the algorithm to simulation right type I censored data

4.4.1 The reliability distribution

The probability of a value - random number generate - falling between a region $(x, +\infty)$ is:

$$P(x_1 > X) = \int_{x_1}^{\infty} f(x)dx$$

Which can see as the same of the definition of the function $R(t)$ - reliability

$$R(x) = \int_{x_1}^{\infty} f(x)dx$$

All random number generates that fall in region $(x, +\infty)$ are the censored data; It's easy to achieve this relation between the reliability and the number $C\%$ of censored data with the expression:

$$R(t_c) = \int_{t_c}^{\infty} f(t)dt = C \tag{4.1}$$

In conclusion, define first the $C\%$ of censored data and then solve the equation 4.1 in order to time t_c and the result correspond to the value in which the function have $C\%$ censored data. In the field of reliability there is some distribution that are more often used. First of them the *Weibull* distribution, and then *exponential, log-normal, gamma* and finally the *normal* distribution. For each of them, it's derive the formulation from specific *PDF* to determine T_c.

4.4.1.1 Weibull distribution

There are several authors that explain very well the Weibull distribution, as can see in Horst (2009) and Murthy et al. (2004).

From the section 2.2.4, the Weibull density function is given by:

$$f(x, \eta, \beta) = \frac{\beta}{\eta^{\beta}} t^{\beta-1} e^{-(\frac{t}{\eta})^{\beta}} \; with \; t \in \mathbf{R}^{+}$$

And the corresponding reliability function is

$$R(t) = e^{-\left(\frac{t}{\eta}\right)^{\beta}}$$

The shape parameter β is a non-dimensional parameter and reflect the type of failure mode.

To have $C\%$ of censored data, it's the same to equal the expression of reliability:

$$R(t_c) = \int_{t_c}^{\infty} f(x)dx = e^{-\left(\frac{t}{\eta}\right)^{\beta}} = C$$

and solve the equation in order of t_c, results:

$$t_c = \eta * (-\log(C))^{\frac{1}{\beta}}$$

that give us the time censoring when we have the $C\%$ of censored data required.

4.4.1.2 Gamma distribution

The gamma distribution represents the sum of n exponentially distributed random variables as is explain in Birolini (2017).

For a random variable, X, and symbolically write, $X \sim G(\alpha, \lambda)$ the probability density function (see in section 2.2.3) is given by:

$$f(x, \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{(\alpha-1)} e^{-\lambda x}, \quad x > 0 \quad \alpha, \lambda > 0$$

Where λ is the failure rate (complete failures) is calculate from $\lambda(t) = f(t)/(1 - F(t))$.

The number of survival items it's the same to have $C\%$ of censored data and can be achieve with the function of reliability:

$$R(t) = e^{-\left(\frac{t}{\eta}\right)^\beta} = C$$

$$R(t_c) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_{t_c}^{+\infty} t^{(\alpha-1)} e^{-\alpha t} dt = C$$

The expression don't have analytic solution, it's necessary to use a numerical resolution method.

4.4.1.3 Log-normal distribution

The log-normal is a versatile distribution and often a better fit to reliability data, such as for populations with wear-out characteristics.

The log-normal pdf is (see section 2.2.6):

$$f(x) = \frac{1}{\sqrt{2\pi} x \sigma^2} e^{-\frac{1}{2\sigma^2}(lnx - \mu)^2}.$$

where μ and σ are the mean and the standard deviation of the ln data.

The log-normal distribution describe reliability of items in which the hazard rate increases from $x{=}0$ to a maximum and then decreases. In this case the reliability functions is:

$$R(t_c) = \int_{t_c}^{+\infty} \frac{1}{\sqrt{2\pi} t \sigma^2} e^{-\frac{1}{2\sigma^2}(\ln t - \mu)^2 dt} = C$$

The expression don't have analytic resolution, it's necessary to use a numerical resolution method.

4.4.1.4 Exponential distribution

The exponential distribution is characterized for the hazard rate be constant. It is considered a homogeneous poisson process. A continuous random variable having probability density function:

$$f(x|\lambda) = \begin{cases} 0 & (x \leq 0), \\ \lambda e^{-\lambda x} & (x > 0), \end{cases}$$

for some $\lambda > 0$ is said to be an exponential random variable with parameter λ and symbolically express, $X \sim Ex(\lambda)$. To have $C\%$ of censored data, it's the same to equal the expression of reliability:

$$R(t_c) = \int_{t_c}^{\infty} \lambda e^{-\lambda t} = e^{-\lambda t_c} = C$$

and solve the equation in order of t_c, results:

$$t_c = -\frac{\ln(C)}{\lambda}$$

that give the time censoring with the $C\%$ of data censored required.

4.4.1.5 Normal distribution

A random variable X has a standard distribution with parameters μ and σ^2 with a density function of the form:

$$f(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

It is written symbolically, $X \sim N(\mu, \sigma^2)$

The parameters of the normal distribution are represented by μ and σ^2 because they correspond respectively to the mean and variance of the random variable.

To have $C\%$ of censored data, it's the same to equal the expression of reliability, and the reliability function is defined by the integral:

$$R(t_c|\mu, \sigma^2) = \int_{t_c}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\}} dt = C$$

for which no analytical solution is known. The values of the t_c time censoring must then be calculated using numerical analysis methods.

4.4.2 The simulation study

For this study and analysis, the distributions used are the most suitable in reliability and maintainability: *Weibull, normal, gamma, log-normal and exponential.* For each distribution it's calculate the censorship time t_c to different values. In great number of reliability

studies they chose the same, or similar, range of values *C% - percentages of censorship*: - 5%, 10%, 20% or 30%. The number of cycles simulations $\boldsymbol{M}$ is **1000 times** and two random sample size of *n=50* and *n=1000*, to take care of medium and large data sets. The algorithm and the program to made this simulation are refer in section 4.4

4.4.2.1 Results from Weibull distribution

The study from Weibull distribution perform an analysis for the shape factor β with a range from *0.5, 1, 1.5, 2, 3 and 5*, which are very illustrative of the shape factor β influence; the scale factor used is $\eta = 1$. The resume of the study are in two tables that summarize the analysis. The first table is the simulation made with sample $n = 50$ and the second is with sample $n = 1000$.

	5			10			**20**			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\beta_{0.5}$	7.67	2.87	14.5	5.18	1.96	2.3	2.60	0.88	0.5	1.47	0.51	1.3
β_1	2.64	0.51	11.8	2.23	0.42	3.2	1.58	0.28	1.6	1.22	0.22	0.9
$\beta_{1.5}$	1.88	0.29	9.5	1.73	0.20	0.6	1.33	0.15	3.1	1.11	0.15	1.8
β_2	1.61	0.17	6.9	1.51	0.13	0.8	1.24	0.11	2.4	1.09	0.09	0.8
β_3	1.38	0.08	4.0	1.29	0.08	2.6	1.16	0.08	1.0	1.06	0.06	0.5
β_5	1.20	0.05	3.3	1.17	0.04	0.9	1.09	0.04	0.7	1.03	0.03	0.6

Table 4.1: Simulation right type I , Weibull $(\beta, C\%)$, $\eta = 1$, $n = 50$

	5			10			**20**			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\beta_{0.5}$	8.93	0.81	0.5	5.27	0.44	0.7	2.59	0.21	0.2	1.46	0.12	0.4
β_1	2.99	0.14	0.2	2.30	0.10	0.2	1.61	0.06	0.1	1.20	0.05	0.2
$\beta_{1.5}$	2.08	0.06	0.1	1.74	0.05	0.2	1.37	0.04	0.0	1.13	0.03	0.1
β_2	1.73	0.04	0.3	1.52	0.03	0.1	1.27	0.02	0.2	1.10	0.02	0.2
β_3	1.44	0.02	0.2	1.32	0.02	0.1	1.17	0.02	0.1	1.06	0.01	0.0
β_5	1.24	0.01	0.1	1.18	0.01	0.0	1.10	0.01	0.1	1.04	0.01	0.1

Table 4.2: Simulation right type I, Weibull $(\beta, C\%)$, $\eta = 1$, $n = 1000$

The results of the Weibull distribution are very interesting. With a sample of $n = 50$ and a cycle of simulations $M = 1000$, the standard deviation is higher when $\beta = 0.5$ and then slowly decrease until $\beta = 5$. The *percentage error PE* it's much higher when the *C%=5%* and then the value goes down when the shape factor increase. There is a different behaviour in *C%=20%*, in this case the standard deviation and the *percentage error PE* is higher when $\beta = 1.5$, this could have some explanation possible or relationship with the transition from exponential shape to normal shape.

The table 4.2 show an simulation to a sample of *1000* and in this case the *standard deviation and PE* is smaller than in the case when sample is *50* from table 4.1. There is

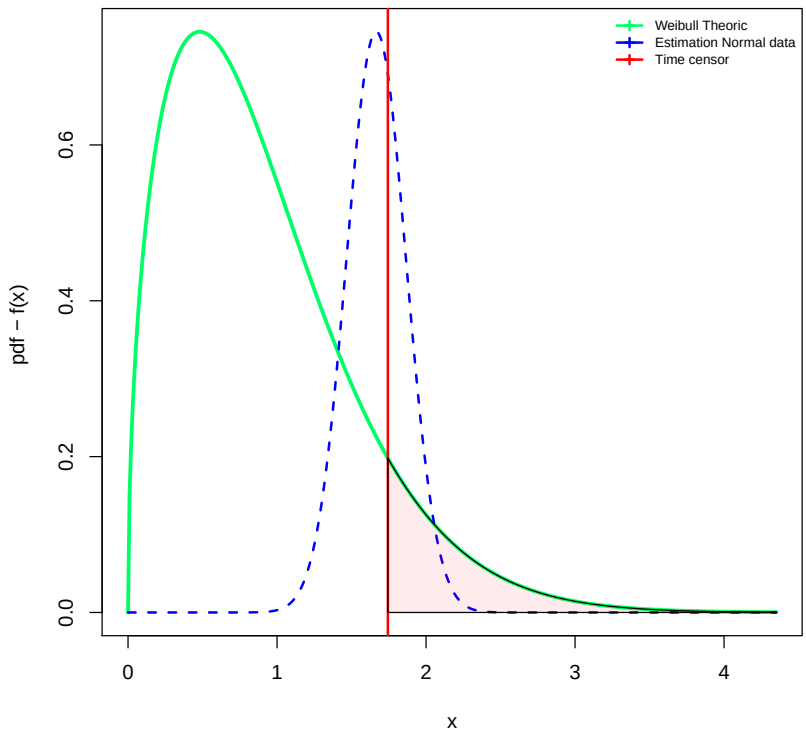


Figure 4.9: Simulation right type I, Weibull - $\beta = 1.5$, $C = 10\%$, $n = 50$

a small decrease in dispersion and error as the shape factor increases. Finally, the Weibull distribution, use well these algorithm, but it's need to have caution and choose a correct number of sampling in order to give more accuracy to the simulation study.

In figure 4.9 can see graphically the dispersion, the bias and the mean of simulation and compare theoretical curve with the normal function from estimation data and the fixed value t_c calculated.

In annex B it's possible to see an exhaust number of figures for all statistical distribution and for different values of sampling, shape factor and $\%C$ percentage of data censored. To all simulation it can be noted that the error is less than 1% , which is very small and even the dispersion itself is very small, as can be graphically seen in all figures.

4.4.2.2 Results from gamma distribution

The results with the gamma distribution are quite similar to the Weibull function. In this distribution the variation of the shape factor α don't have so influenced like in Weibull distribution. In this case, the values to the standard deviation from sampling *50* and sampling *1000* is very similar. The standard deviation is sightly increase when the failure rate increase but the absolute values is small. The percentage error are very different from

sampling 50 and sampling 1000. In sampling *50* the *PE* is very high and in sampling *1000* is small and quite stable.

The algorithm develop to gamma distribution are very suitable, but it's need to have caution and choose a great number of sampling in order to give more accuracy to the simulation study.

	5			10			**20**			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\alpha_{0.5}$	1.61	0.41	16.1	1.29	0.32	5.0	0.81	0.19	1.6	0.53	0.16	1.4
α_1	2.56	0.47	14.6	2.18	0.41	5.5	1.56	0.26	2.8	1.19	0.22	1.4
$\alpha_{1.5}$	3.55	0.58	9.1	3.01	0.44	3.7	2.23	0.34	4.0	1.82	0.26	0.5
α_2	4.27	0.61	10.0	3.79	0.54	2.5	2.88	0.37	3.8	2.43	0.29	0.5
α_3	5.83	0.74	7.4	5.25	0.64	1.3	4.24	0.41	1.0	3.64	0.40	0.8
α_5	8.64	0.88	5.6	7.88	0.62	1.5	6.67	0.51	0.7	5.82	0.50	1.1

Table 4.3: Simulation right type I, Gamma $(\alpha, C\%)$, $\lambda = 1$, $n = 50$

	5			10			**20**			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\alpha_{0.5}$	1.91	0.11	0.4	1.35	0.08	0.3	0.82	0.05	0.2	0.54	0.03	0.3
α_1	3.00	0.14	0.0	2.30	0.09	0.0	1.60	0.06	0.3	1.20	0.05	0.2
$\alpha_{1.5}$	3.90	0.16	0.1	3.12	0.11	0.3	2.32	0.08	0.1	1.83	0.06	0.3
α_2	4.74	0.16	0.1	3.88	0.12	0.2	2.99	0.08	0.1	2.43	0.07	0.2
α_3	6.28	0.19	0.3	5.32	0.14	0.1	4.28	0.10	0.0	3.61	0.08	0.0
α_5	9.14	0.22	0.1	7.98	0.16	0.1	6.72	0.12	0.1	5.88	0.11	0.1

Table 4.4: Simulation right type I, Gamma $(\alpha, C\%)$, $\lambda = 1$, $n = 1000$

4.4.2.3 Results from Log-normal distribution

The Log-normal distribution have an behaviour very interesting. When increase the σ standard deviation the *PE* percentage error and the standard deviation increase. There is some difference in the values from a simulation that have sample of *50* and a sample of *1000*. The value is much higher and the bias and Percentage error is much higher in the sample of *50*.

In the case of the log-normal distribution standard deviation parameter have influence when increase the value.

	5			10			**20**			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\sigma_{0.3}$	4.23	0.34	5.0	3.98	0.29	0.4	3.47	0.22	0.9	3.17	0.17	0.4
$\sigma_{0.5}$	5.70	0.80	7.8	5.04	0.57	2.3	4.08	0.38	1.4	3.50	0.32	1.0
$\sigma_{0.7}$	8.01	1.41	6.8	6.54	0.97	1.9	4.78	0.77	2.4	3.97	0.49	1.2
σ_{1}	11.91	2.93	15.4	9.28	2.06	5.2	6.16	1.30	2.3	4.56	0.85	0.7
$\sigma_{1.3}$	18.86	7.16	18.2	13.69	4.11	4.8	7.72	2.09	4.9	5.29	1.28	1.6
$\sigma_{1.7}$	32.92	14.02	26.1	22.05	8.68	8.2	10.70	3.71	5.9	7.05	2.06	6.3

Table 4.5: Simulation right type I, Log-normal $(\sigma, C\%)$, $\mu = 1$, $n = 50$

	5			10			**20**			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\sigma_{0.3}$	4.45	0.09	0.1	3.99	0.06	0.0	3.50	0.05	0.0	3.18	0.04	0.0
$\sigma_{0.5}$	6.17	0.21	0.3	5.15	0.15	0.1	4.14	0.09	0.0	3.53	0.07	0.0
$\sigma_{0.7}$	8.58	0.41	0.2	6.67	0.24	0.0	4.90	0.15	0.1	3.92	0.11	0.1
σ_{1}	14.06	0.98	0.1	9.77	0.52	0.3	6.30	0.28	0.2	4.58	0.20	0.2
$\sigma_{1.3}$	22.98	2.02	0.4	14.39	1.00	0.0	8.10	0.47	0.2	5.37	0.29	0.0
$\sigma_{1.7}$	44.16	4.94	0.8	23.93	2.16	0.4	11.43	0.89	0.5	6.61	0.47	0.3

Table 4.6: Simulation right type I, Log-normal $(\sigma, C\%)$, $\mu = 1$, $n = 1000$

4.4.2.4 Results from exponential distribution

To simulation exponential distribution the hazard rate change from *0.5 to 5*. The standard deviation and the PE is much higher in the simulations with sample *50*. After some analysis, the conclusion is that the exponential distribution is stable and more robust when the sampling is higher. The bias is reduced when the *C%* and the hazard rate increase.

	5			10			**20**			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\lambda_{0.5}$	5.44	0.90	9.2	4.40	0.87	4.5	3.17	0.53	1.4	2.39	0.46	0.6
λ_{1}	2.74	0.55	8.7	2.25	0.43	2.5	1.59	0.30	1.2	1.16	0.20	3.4
$\lambda_{1.5}$	1.81	0.35	9.1	1.48	0.27	3.8	1.03	0.17	3.6	0.81	0.13	0.5
λ_{2}	1.36	0.25	9.0	1.13	0.20	1.9	0.76	0.13	5.2	0.60	0.11	0.4
λ_{3}	0.89	0.14	10.9	0.74	0.14	3.3	0.52	0.09	2.9	0.40	0.08	1.2
λ_{5}	0.53	0.09	11.8	0.44	0.10	4.2	0.32	0.05	1.7	0.24	0.04	1.3

Table 4.7: Simulation right type I, Exponential $(\lambda, C\%)$, $n = 50$

	5			10			**20**			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\lambda_{0.5}$	5.96	0.27	0.6	4.59	0.20	0.3	3.21	0.13	0.3	2.40	0.10	0.2
λ_1	2.99	0.13	0.2	2.30	0.09	0.0	1.61	0.06	0.1	1.20	0.05	0.2
$\lambda_{1.5}$	1.99	0.09	0.2	1.53	0.06	0.0	1.07	0.04	0.2	0.80	0.03	0.1
λ_2	1.49	0.07	0.3	1.15	0.05	0.4	0.81	0.03	0.1	0.60	0.02	0.1
λ_3	1.00	0.05	0.1	0.77	0.03	0.3	0.54	0.02	0.1	0.40	0.02	0.1
λ_5	0.60	0.03	0.1	0.46	0.02	0.1	0.32	0.01	0.2	0.24	0.01	0.0

Table 4.8: Simulation right type I, Exponential $(\lambda, C\%)$, $n = 1000$

The exponential distribution can be used in the algorithm developed, but it's need to be very carefully and choose a right number of sample in order to give more precision to the simulation study.

4.4.2.5 Results from normal distribution

The simulations with normal distribution use the mean $\mu = 1$ and the standard deviation range from 0.5 to 5. Comparing the two table is quite easy to conclude that the bias and the PE is much higher in the simulations with sample *50*. The normal distribution, when the sampling is higher have their values consistent, low bias and reliable results.

	5			10			**20**			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\sigma_{0.5}$	1.74	0.14	4.7	1.61	0.12	2.2	1.40	0.11	1.7	1.24	0.10	1.6
σ_1	2.46	0.26	7.0	2.19	0.24	3.8	1.82	0.21	1.3	1.49	0.20	2.4
$\sigma_{1.5}$	3.14	0.41	9.5	2.77	0.35	5.1	2.21	0.29	2.5	1.69	0.27	5.5
σ_2	3.85	0.50	10.3	3.48	0.49	2.3	2.64	0.40	1.8	1.90	0.31	7.2
σ_3	5.34	0.72	10.0	4.60	0.69	5.0	3.42	0.59	3.1	2.52	0.53	2.0
σ_5	8.55	1.40	7.3	7.22	1.03	2.5	5.09	1.04	2.2	3.56	0.95	1.8

Table 4.9: Simulation right type I, Normal $(\sigma, C\%)$, $\mu = 1$, $n = 50$

	5			10			**20**			30		
	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\sigma_{0.5}$	1.82	0.03	0.1	1.64	0.03	0.1	1.42	0.02	0.1	1.26	0.02	0.1
σ_1	2.64	0.07	0.3	2.28	0.05	0.2	1.84	0.05	0.1	1.52	0.04	0.1
$\sigma_{1.5}$	3.46	0.10	0.2	2.92	0.08	0.1	2.26	0.07	0.1	1.79	0.06	0.1
σ_2	4.28	0.13	0.1	3.56	0.11	0.1	2.68	0.09	0.1	2.04	0.09	0.3
σ_3	5.92	0.20	0.3	4.84	0.17	0.0	3.52	0.14	0.1	2.57	0.12	0.2
σ_5	9.20	0.34	0.2	7.39	0.27	0.3	5.20	0.22	0.2	3.61	0.21	0.3

Table 4.10: Simulation right type I, Normal $(\sigma, C\%)$, $\mu = 1$, $n = 1000$

The algorithm developed can be used with normal distribution, but need to be very attention to the results and the behaviour of the random number generator in order to have the results reliable and with accuracy.

4.5 Simulation and test randomness of data censored

This section intends to study in depth the construction of an algorithm that does the simulation of the fixed right type I censored data. That is, the data censored at the end of the study because the items continues to work or has survived at the defined time (see section 4.1.3)

The definition of censorship time have been explained in the previous section. In this section, the algorithms for simulating censoring data are developed and the methodology to evaluate if the this algorithms produce data that can be considered random or not.

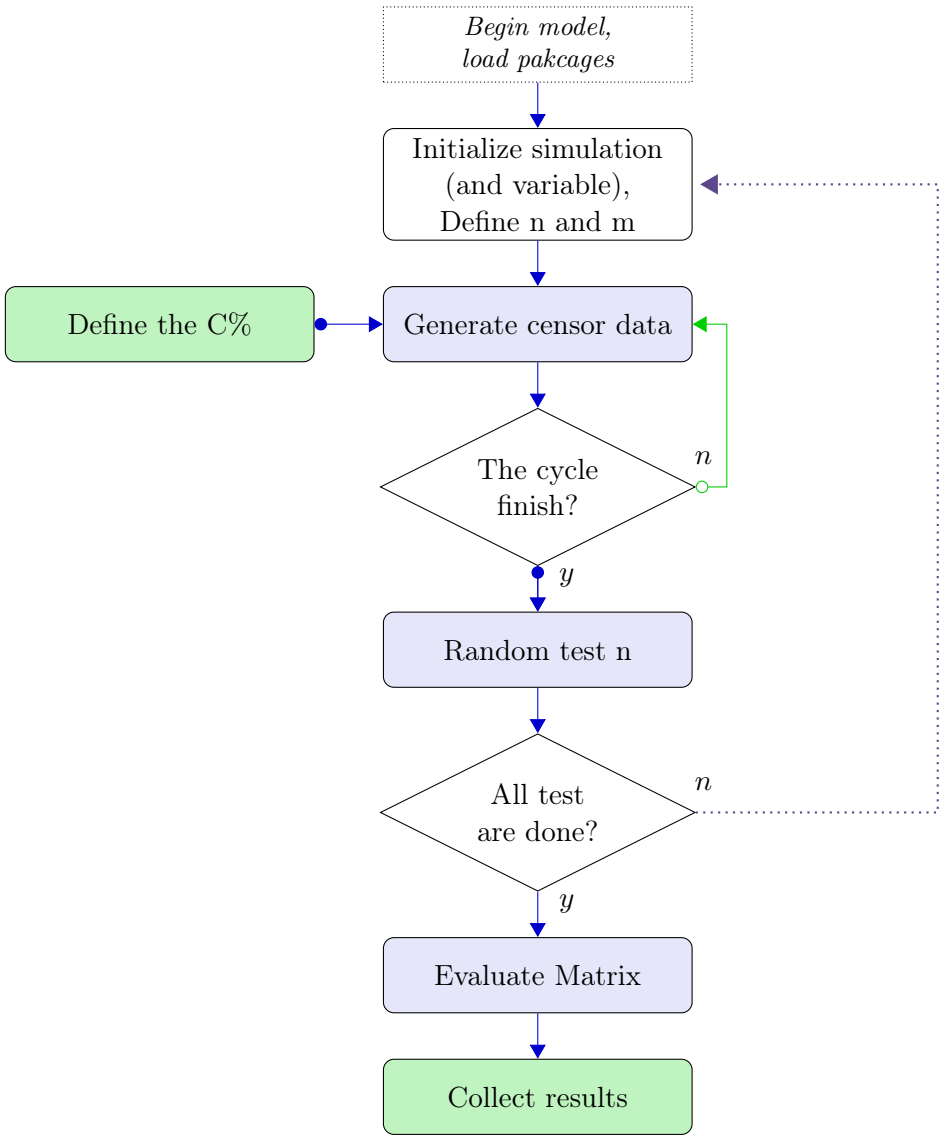


Figure 4.10: Methodology of simulation and randomness test of data censored

The methodology developed starts to generate random numbers based on the uniform distribution and then converting to the chosen distribution, using the defined parameters. A sample number n to generate t_i random times for each simulation is chosen, then the *five* random tests are made and for each test is calculated the p-value (for the chosen significance level - $\alpha = 0.05$ and $\alpha = 0.1$).

The cycle of simulation is *M=1000 times* and an average is made for each cycle of simulation. This routine is made for *five* parameters from Weibull, normal, gamma, log-normal and exponential distribution.

For each cycle of different parameter and different censorship percentage, it's give a grade from *1* to *3* regarding the evaluation matrix see in table 4.11 and the scale have the follow explanation:

- **3** points correspond to a percentage that (100-90)% tests are *iid* in the and the algorithm can be considerer *good/accepted*,
- **2** points correspond to a percentage in the interval (90-75)% that the results of the tests give *iid* and the algorithm can be consider *admissible/accepted*.
- **1** point correspond to a percentage (75-0)% and the algorithm is *bad/not acceptable*.

Interval %	**Description**	**Grade**	**Status**
(100%-90%)	good / accepted	3	OK
(90%-75%)	reasonable / accepted	2	Adm
(75%-0%)	bad / not acceptable	1	NOK

Table 4.11: Matrix Evaluation for Randomness

The study make an overall assessment that does the classification using the evaluation matrix. This work is made for each distribution, comparing with the different parameters and different censorship values.

This methodology can be a very good tool to use and classified the algorithms that generate the censored data.

4.6 Simulation and test of censored data fixed right type I

The general algorithm to fixed right type-I censor data have these steps:

- STEP 1 Define initial parameters and values to simulation (=0)
- STEP 2 Calculate censoring-time - T_C (with parameters of distribution and number of data censor C%)
- STEP 3 Generate the vector Y that represent n random times t_i from distribution model (defining a function)

- STEP 4 Defining the times censored and return the final times with vector T_{cens}
- STEP 5 Apply to vector T_{cens} the five random test and save to vector $pval_{testx}$ the result
- STEP 6 From vector $pval_{testx}$ calculate the significance and the accept or rejection sample from defining α and save the result in vector r_{testx}
- STEP 7 Repeat M times from step 3 to 6
- STEP 8 Calculate the number of samples that accept the hypothesis h_0
- STEP 9 Calculate and classification from different shapes and samples if the model is OK, Adm. or NOK, using the evaluation matrix
- STEP 10 Plot and save the results to table

The program made to Weibull distribution is then extrapolate to different distribution. The first part of the code in R software is :

```
1. Generate function
#  Right fixed type I weibull data censored.
censtypeI <- function(n,shape,scale,censnumber){
  T <- rweibull(n, shape, scale) # generate the random number
  # time theoretical censor from weibull
  censoring_time <- scale*(-log(censnumber))^(1/shape)
    for (j in 1:n)
        {T_cens[j] <- min(censoring_time, T[j]) }
return(T_cens)  }
```

Algorithm 2: Program in R from Weibull distribution - define generate function

The program made in R, in order to optimize and reuse, used functions that could be invoked and integrated in the routine. This cycle/loop became the main nucleus of the simulation: define the function of random number generation, the function of time of censorship and also the function of final vector with the censored data. The structure of the algorithm is the same for all distribution and only change the mathematical model of each distribution.

The step *5* of the algorithm is to make the tests of randomness. To make the randomness test the package *randtests* developed by Caeiro and Mateus (2014) is used. There are other packages available to do these tests; this one seemed to be the most adequate, robust and simple for the purpose. Make the programs for each random test from the beginning will be wasting time and will delay all research work.

Finally, step *9* define the routines to classification of results as well the overall grades and then save and plot the results in tables and figures.

```
2. Random test
## ######## Funcão de teste de aleatoriedade #####
Randomtest <- function(m,n,shape,alpha,cen){
# Initial values form vector that colect the p-values
pval_Trun =0; pval_TRank=0;pval_Tbartels=0
pval_Tturning=0; pval_TKs=0
# beginning of the cycle (by m)
for (i in 1:m){
  y <- censtypeI(n,shape,scale,cen)
    # Run test
        rt <- RunsTest(y)
        pval_Trun[i] <- rt$p.value
        if (pval_Trun[i]== "NaN") {pval_Trun[i]=0} else { }
    # Rank test
        rt <- rank.test(y)
        pval_TRank[i] <- rt$p.value
    #Turning point test
        rt <- turning.point.test(y)
        pval_Tturning[i] <- rt$p.value
    # BartelsRank Test
        rt <- BartelsRankTest(y, alternative="left.sided",
                              pvalue="normal")
        pval_Tbartels[i] <- rt$p.value
    #Komogorov-Smirnov Test
        rt <- ks.test(y, "pweibull", shape,
                      scale,alternative = "l")
        pval_TKs[i] <- rt$p.value
    }
testresult <- data.frame(pval_Trun,pval_TRank,
                         pval_Tturning,pval_Tbartels,pval_TKs)
return(testresult) }
```

Algorithm 3: Program in R from Weibull distribution - Tests for randomness

To illustrate and compare the methods as described above, a random sample of different sizes, *n=10, 100, 500, 1000,10000* to take care of small, medium and large data sets.

The simulation complete cycle/loop make for each distribution and for each percentages of censored data: *5%, 10%, 20% and 30%* and for the level of significance $\alpha = 0.05$ and $\alpha = 0.01$.

All the results are in the appendix C, here, only show the simulation and results made to $\alpha = 0.05$ and to Weibull, Gamma and Normal distribution. The rest of the simulations and tables are in appendix C and permit to complete, refine and give strength and robustness to the final conclusions.

4.6.1 Results from Weibull distribution

The algorithm described in section 4.6 is applied for the Weibull distribution with the variable parameters β and $C\%$. And the β - shape factor range from *0.5, 1, 1.5, 2, 3 and 5* and censored data parameter from *5%, 10%, 20% and 30%*. The fixed parameter are $n = 100$, $M = 1000$ and scale factor $\mu = 1$.

To the level of significance $\alpha = 0.05$ and $\alpha = 0.01$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\beta_{0.5}$	0.97	0	0	0.98	0.96	0.96	0.91	0.92	0.97	0.96	0.95	0.61	0.5	0.95	0.96	0.95	0.07	0.02	0.97	0.95
β_1	0.97	0	0	0.98	0.96	0.95	0.93	0.92	0.95	0.95	0.95	0.6	0.52	0.94	0.94	0.95	0.1	0.02	0.96	0.95
$\beta_{1.5}$	0.96	0	0	0.98	0.95	0.96	0.92	0.93	0.96	0.95	0.95	0.62	0.52	0.96	0.94	0.94	0.09	0	0.95	0.95
β_2	0.97	0	0	0.98	0.94	0.95	0.93	0.92	0.96	0.95	0.96	0.63	0.51	0.96	0.96	0.96	0.09	0.02	0.95	0.95
β_3	0.97	0	0	0.98	0.95	0.95	0.93	0.92	0.96	0.96	0.96	0.63	0.54	0.96	0.96	0.95	0.1	0.01	0.96	0.95
β_5	0.96	0	0	0.96	0.93	0.94	0.92	0.91	0.95	0.94	0.95	0.63	0.54	0.95	0.94	0.96	0.09	0.02	0.96	0.95

Table 4.12: Test RNG, fixed right type I, Weibull $(\beta, C\%)$, $\alpha = 0.05$, $\eta = 1$, $n = 100$

The simulation and the random test to data censored fixed right type I with Weibull distribution give some interest results. There is no significant differences when the shape factor β increase from *0.5 to 5* to all tests. There is a smooth decrease when the $C\%$ of censored data increase from *5% to 30%* and this is more evident to $\alpha = 0.01$ (see annex C). To each random test the behaviour is the same to $\alpha = 0.05$ and $\alpha = 0.01$ (see annex C).

The R_u - *run test* accept in all situation the hypotheses of *iid* of the samples and the same is to *Bartels* and *Kolmogorov-Smirnov* tests.

The R_a - *rank test* start to reject in all simulations in the first situation $C_{5\%}$ but accept partial the randomness when have $C_{10\%}$ and $C_{20\%}$ and completely reject to $C_{30\%}$. Probably the algorithm will fail for complete data and when the sample are very censored.

The turning-point test have the same behaviour of the rank test. This exceptional results are very difficult to interpret, but it's fundamental to experiment and try to understand what happens. In this case the results can suggest that probably in the first situation $C_{5\%}$ (near of complete data) the test indicate uniformity and in the second situation $C_{30\%}$ the *iid* of samples is reject.

4.6.2 Results from Gamma distribution

The study performed an analysis for Gamma distribution and the shape factor α ranging from *0.5, 1, 1.5, 2, 3 and 5*, which are very illustrative of the α variation.

The Gamma distribution have a similar simulation and results in random tests to censored data fixed right type I as Weibull distribution. There is no significant change when the shape factor α increase from *0.5 to 5* to all tests. There is a trend of decrease

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\alpha_{0.5}$	0.98	0	0	0.98	0.95	0.96	0.92	0.93	0.95	0.95	0.96	0.63	0.54	0.95	0.95	0.95	0.08	0.01	0.95	0.94
α_1	0.97	0	0	0.97	0.94	0.96	0.92	0.92	0.96	0.95	0.94	0.62	0.52	0.96	0.96	0.94	0.09	0.02	0.95	0.95
$\alpha_{1.5}$	0.97	0	0	0.98	0.95	0.95	0.93	0.91	0.96	0.96	0.95	0.64	0.55	0.94	0.96	0.95	0.08	0.01	0.94	0.95
α_2	0.98	0	0	0.98	0.95	0.95	0.92	0.92	0.95	0.96	0.95	0.61	0.54	0.95	0.95	0.96	0.11	0.02	0.96	0.94
α_3	0.97	0	0	0.98	0.95	0.96	0.92	0.93	0.96	0.95	0.96	0.63	0.48	0.95	0.96	0.96	0.08	0.01	0.96	0.96
α_5	0.97	0	0	0.97	0.94	0.95	0.92	0.94	0.94	0.96	0.96	0.63	0.52	0.95	0.96	0.96	0.09	0.01	0.96	0.95

Table 4.13: Test RNG, fixed right type I, Gamma $(\alpha, C\%)$, $\alpha = 0.05$, $\lambda = 10$, $n = 100$

when the $C\%$ of censored data increase from *5% to 30%* and this is more evident to $\alpha = 0.01$ (see annex C).

The *run test* accept in all cases the hypotheses of *iid* of the samples and to Bartell and Kolmogorov-Smirnov tests is the same.

The rank test begin to reject in all simulations in the first situation $C_{5\%}$ but accept partial the randomness in $C_{10\%}$ and $C_{20\%}$ and completely reject to $C_{30\%}$. Probably the algorithm will fail for complete data and when the sample are very censored.

The turning-point test have the same behaviour of the rank test.

The evolution and the explanation is in line with the same conclusion made to Weibull distribution.

4.6.3 Results from Normal distribution

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\sigma_{0.5}$	0	0	0	0.98	0.94	0	0.92	0.92	0.95	0.94	0	0.62	0.55	0.96	0.96	0	0.09	0.02	0.94	0.94
σ_1	0	0	0	0.98	0.94	0	0.93	0.92	0.96	0.95	0	0.66	0.55	0.97	0.96	0	0.08	0.01	0.97	0.95
$\sigma_{1.5}$	0	0	0	0.97	0.94	0	0.92	0.92	0.95	0.94	0	0.63	0.5	0.95	0.96	0	0.09	0.02	0.95	0.96
σ_2	0	0	0	0.97	0.95	0	0.93	0.93	0.94	0.94	0	0.64	0.49	0.94	0.95	0	0.1	0.01	0.94	0.95
σ_3	0	0	0	0.97	0.95	0	0.94	0.92	0.96	0.96	0	0.61	0.51	0.95	0.95	0	0.08	0.02	0.95	0.96
σ_5	0	0	0	0.98	0.94	0	0.94	0.93	0.96	0.94	0	0.65	0.54	0.96	0.95	0	0.1	0.01	0.96	0.93

Table 4.14: Test RNG, fixed right type I, Normal $(\sigma, C\%)$, $\alpha = 0.05$, $\mu = 1$, $n = 100$

The study with Normal distribution did a simulations for the standard deviations σ parameter to varies between *0.5, 1, 1.5, 2, 3 and 5* values that are quite illustrative. The other parameter of distribution is the mean μ and the value is *1*.

The Normal distribution have the standard deviation that change but it's not the same as Weibull or Gamma function that use parameters that change the shape form of the function.

The R_a rank test and the T_u Turning point tests have the same behaviour as Weibull and Gamma distribution. The most positive tests was the Bartels and Kolmogorov-Smirnov tests that accept the *iid* of the sample in all situations. The R_u *run test* is the opposite of the last two test and fails in all situations. Probably this behaviour have some relations directly with the core of the hypotheses test and the use of the normal distribution to calculate the variable of statistics test.

4.6.4 Global results of fixed right type I

The final result for the criteria that are define in the section 4.5 are resume in the table *4.15*.

		$C_{(5\%)}$	$C_{(10\%)}$	$C_{(20\%)}$	$C_{(30\%)}$
Weibull	$\alpha_{0.05}$	Adm	OK	Adm	Adm
	$\alpha_{0.1}$	NOK	Adm	NOK	NOK
normal	$\alpha_{0.05}$	NOK	OK	NOK	NOK
	$\alpha_{0.1}$	NOK	NOK	NOK	NOK
lognormal	$\alpha_{0.05}$	Adm	OK	Adm	Adm
	$\alpha_{0.1}$	NOK	Adm	NOK	NOK
gamma	$\alpha_{0.05}$	Adm	OK	Adm	Adm
	$\alpha_{0.1}$	NOK	Adm	NOK	NOK
exponencial	$\alpha_{0.05}$	NOK	OK	NOK	NOK
	$\alpha_{0.1}$	NOK	NOK	NOK	NOK

Table 4.15: Global Results of Simulation Test - Right Fixed Type I - Shape Factor

The table *4.15* permit some conclusions about the random number generation of data censored fixed right type I: - the most favourable and reliable distribution is the Weibull and Gamma distribution, fortunately is the two most used distribution on reliability. The worst distributions that failed almost in all criteria is the Normal and after that the Exponential distribution; - in the middle there is one distribution - Log-normal.

This results show the importance of doing a set of random test to the algorithm of generating random number with data censored. In some cases, will be necessary to change the algorithm or the parameters or the statistical distribution.

In conclusion, when the researchers needs to generate random number with data censored, the simulation process need to have extra step to confirm and validate the random number generation.

4.7 Simulation and test of censored data right type-II

This section study in depth the construction of an algorithm that does the simulation of right type II censored data, that is, when data is censored at the end of r failures or events.

The methodology starts with random number generation based on uniform distribution, then converting to the chosen distribution, as well as the parameters defined. The analysis will be only made in two distributions, leaving for future work, the other distributions.

In this section, a preliminary study have been done in order to conclude, if it would be interesting to applied the random tests, because the final result of the algorithm of random number generation censored data right type II are a set of number grouped and sorted, and this retired all randomness and efficiency to the random tests. But, after a deep analysis and some experiences with different algorithms and random test, its verified is real important to test the random of the original data, i.e., data without sorting and after removing the censored values, or by other words, data that remain with the value above T_c (n of chosen failure).

The five randomness tests are applied and for each of them found the *p-value* (for the significance level chosen - $\alpha = 0.05$ and $\alpha = 0.01$), then made *M= 1000* simulation and average the *p-values* of each distribution. This routine is made for different shape parameters (five) of two distribution and for six sample sizes.

For each cycle of different α and different censorship percentage $C\%$, evaluation is made according to the matrix 4.11 in order to classified the results regarding the overall results: *3* if is *OK*, *2* if it is *Adm* - admissible or *1* if is not acceptable - *NOK*.

The study did an overall assessment of the evaluation for two distribution, comparing with the different level of significance α and different C% of censored data .

This methodology pretends to examine and conclude in a reasonable way the validation of the use of the algorithms to generate the right type II censored data.

The general algorithm to right type-II censored data have these steps:

- STEP 1 Define initial parameters and values to simulation (=0)
- STEP 2 Calculate censoring-time - T_C (with parameters of distribution and number of data censor C)
- STEP 3 Generate the vector Y that represent n random times t_i from distribution model (defining a function)
- STEP 4 Defining the times censored and return the final times with vector T_{cens}
- STEP 5 Apply to vector T_{cens} the five random test and save to vector $pval_{testx}$ the result
- STEP 6 From vector $pval_{testx}$ calculate the significance and the accept or rejection sample from defining α and save the result in vector r_{testx}

- STEP 7 Repeat M times from step 3 to 6
- STEP 8 Calculate the number of accept percentage of samples
- STEP 9 Calculate and classification from different shapes and samples if the model is OK, Adm. or NOK
- STEP 10 Plot and save the results to table

An example of program from Weibull distribution, made in *R* software:

```
censtypeII <- function(n,shape,scale,cens_failure){
  # generation ti and define parameters
  r <- n*(1-cens_failure)
  failure_number <- round(r,0)
  T <- rweibull(n, shape, scale)
  #calculate the vectors of data censored
  T_order <- sort(T) #vector sort ascending numbers
  T_cens <- T_order[1:failure_number]
  T_max <- T_order[failure_number]
  T_rnd <-  T[T<=T_max]
  return(T_rnd)
}
y <- censtypeII(100,0.5,10,0.05)
```

Algorithm 4: Program in R from Weibull distribution - define generate function

The simulation was made in the two distributions for the following percentages of censored data: 5%, 10%, 20% and 30% and for the level of significance α with values of *0.05* and *0.01*.

4.7.1 Results from Weibull distribution

Using the methodology described for the Weibull distribution, the analysis is for the shape factor range from *0.5, 1, 1.5, 2, 3 and 5*, which are very illustrative of the β variation and used as scale factor μ the value *1*.

The simulation and the random test to right type II censored data with Weibull distribution give some interest results. To the level of significance $\alpha = 0.05$ and $\alpha = 0.01$ there is some little difference, but not change the rejection or acceptance overall of the test.

When the shape factor β increase from *0.5 to 5* the values decrease to Rank and Bartels test. When the $C\%$ of censored data is 5% the results are very difficult to understand, because the Run test give accept *iid* to all shapes factor β with the value 1, and the Turning point and Kolmogorov-Smirnov tests reject the hypotheses of *iid*.

This situation can appoint to a conclusion that the algorithm of complete data or censored type II with 5% is probably, in most the cases, not random with a strong uniformity of data.

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\beta_{0.5}$	1	0.1	0	0.09	0	1	0.96	0.85	0.96	0.94	1	0.94	0.17	0.94	0.9	1	0.83	0	0.89	0.56
β_1	1	0.08	0	0.08	0	1	0.95	0.87	0.95	0.94	1	0.95	0.16	0.94	0.9	1	0.83	0	0.89	0.56
$\beta_{1.5}$	1	0.07	0	0.09	0	1	0.95	0.86	0.95	0.95	1	0.95	0.17	0.95	0.89	1	0.85	0	0.88	0.56
β_2	1	0.09	0	0.08	0	1	0.95	0.86	0.95	0.96	1	0.95	0.16	0.94	0.88	1	0.83	0	0.89	0.54
β_3	1	0.08	0	0.08	0	1	0.96	0.86	0.96	0.96	1	0.93	0.16	0.96	0.89	1	0.83	0	0.9	0.58
β_5	1	0.07	0	0.08	0	1	0.96	0.86	0.96	0.95	1	0.94	0.18	0.95	0.9	1	0.85	0	0.88	0.56

Table 4.16: Test RNG, right type II, Weibull $(\beta, C\%)$, $\alpha = 0.05$, $\eta = 1$, $n = 100$

The run test in all cases accept the hypotheses of *iid* of the samples. Exception to the case of 5% of censored data in which all the tests accept the hypotheses of random numbers and *iid* of the RNG.

4.7.2 Results from Gamma distribution

The study performed an analysis for the shape factor α from *0.5, 1, 1.5, 2, 3 and 5*, which are very illustrative of the α variation. The failure factor λ have a value of *1*.

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\alpha_{0.5}$	1	0.08	0	0.08	0	1	0.94	0.88	0.95	0.95	1	0.95	0.18	0.96	0.9	1	0.83	0	0.88	0.57
α_1	1	0.09	0	0.08	0	1	0.95	0.86	0.95	0.96	1	0.94	0.16	0.95	0.89	1	0.84	0	0.88	0.55
$\alpha_{1.5}$	1	0.07	0	0.09	0	1	0.95	0.87	0.95	0.95	1	0.93	0.18	0.94	0.9	1	0.84	0	0.9	0.55
α_2	1	0.09	0	0.08	0	1	0.95	0.88	0.95	0.94	1	0.94	0.17	0.94	0.9	1	0.84	0	0.9	0.56
α_3	1	0.08	0	0.08	0	1	0.94	0.87	0.95	0.95	1	0.94	0.18	0.95	0.89	1	0.85	0	0.9	0.54
α_5	1	0.09	0	0.08	0	1	0.96	0.86	0.95	0.95	1	0.95	0.17	0.95	0.9	1	0.85	0	0.89	0.56

Table 4.17: Test RNG, right type II, Gamma $(\alpha, C\%)$, $\alpha = 0.05$, $\lambda = 1$, $n = 100$

The Gamma distribution have the same behaviour of the Weibull distribution data censored right type II.

To all cases the run test accept the hypotheses of *iid* of the sample.

When the shape factor α increase from *0.5 to 5* the values maintains stable to all tests except run test. The interpretation when the $C\%$ of censored data is 5% can be that the algorithm of complete data or censored type II with 5% is probably not random and have uniformity of data.

The Turning point test don't have linear results, because, with 5% of censored data the value is *0*, then increase to up 0.7 when are 10% of censored data and then decrease until arrive at value *0* when censored data is 30%.

4.7.3 Global results of right type II

The final result for the criteria that are define in the section 4.5 is shown in the following resume table:

		$C_{(5\%)}$	$C_{(10\%)}$	$C_{(20\%)}$	$C_{(30\%)}$
	$\alpha_{0.05}$	OK	OK	OK	OK
weibull	$\alpha_{0.1}$	OK	OK	OK	OK
	$\alpha_{0.05}$	OK	OK	OK	OK
gamma	$\alpha_{0.1}$	OK	OK	OK	OK

Table 4.18: Global Results of Simulation Test - Right Type II - Shape Factor

To simplify the research work in this section the most favourable and used distribution have been chosen: the Weibull and Gamma distribution. To the level of significance $\alpha = 0.05$ and the level of $\alpha = 0.01$ and to all different $C\%$ of censored data the algorithm give a very positive results.

The importance of doing a set of random tests to the algorithm of random number generation of right type II censored data, is very useful and give information that confirm if it's necessary change the algorithm, the parameters or the statistical distribution.

This results show the importance of doing a set of random test to the algorithm of generating random number with censored data.

In conclusion, when the researchers needs to generate random number with data censored, the simulation process need to have extra steps to confirm and validate the random number generation.

Chapter 5

Estimation parameters and the method EM

The estimation of parameters of survival functions is one of the most important task in the field of reliability. There is a lot of statistical and mathematical tools and one of the most used is *MLE - Maximum Likelihood Estimation*. Another tool is the Expectation Maximization method (EM) that is an iterative process that can be used to calculate the maximum likelihood estimators when we are in presence of incomplete data or censored data. The EM algorithm is used in a wide range of statistical applications because its formulation reduces the complexity of the estimation problem.

In this chapter the general formulation of censored data using *MLE - Maximum Likelihood Estimation* and EM method was described and some algorithms are developed to compute a simulation using: MLE - Maximum Likelihood Estimation to complete data and to right type I censored data with Weibull distribution in *Python* and *R* software; and Expectation Maximization method to right type I censored data with Weibull distribution in *Python* and *R* software.

5.1 General formulation of censored data

For estimating parameters of survival functions there is a lot of statistical and mathematical tools. One of the classic is the Kaplan-Meier estimator, but sometimes, it's necessary to make more assumptions that allow us to model the data in more detail. Specifying a parametric form for the functions of survival analysis enable to:

- easily compute quantiles of the distribution;
- estimate the expected failure time;
- derive a concise equation for estimating survival functions.

When data is complete the classical MLE - Maximum Likelihood Estimation is enough, but when data are censored it's necessary to assume another conditions, information and parameters.

The likelihood function for right censored has the following form:

$$L(\theta; x, \delta) = \prod_{d \in D} f(x_d) \prod_{r \in R} S(x_r) \tag{5.1}$$

where D is the set of failures times, R is the set of right censored times. For a failure time x_d, $f(x_d)$ is proportional to the probability of observing a failure at time x_d.

For a right censored observed x_r, the only thing that is known is that the real survival time T_r is greater than x_r. Hence, having $P[T_r > x_r] = S(x_r)$, the probability that the real survival time T_r is greater than x_r, for a right censored observation.

The above likelihood can be generalized to the case where there might be any kind of censoring:

$$L(\theta; x, \delta) = \prod_{d \in D} f(x_d) \prod_{r \in R} S(x_r) \prod_{l \in L} S(x_r)[1 - S(x_l)] \prod_{i \in I} [S(U_i) - S(V_i)] \tag{5.2}$$

where L is the set of left censored observations, I is the set of interval censored observations with the only knowledge that the real survival time T_i is in the interval $[U_i, V_i]$. Note that $S(U_i) - S(V_i) = P[U_i \leq T_i \leq V_i]$ is the probability that the real survival time T_i is in $[U_i, V_i]$.

To simplify, some authors express the data censor modelation like this:

- If Y_i is uncensored, the ith subject contributes $f(Y_i)$ to the likelihood
- If Y_i is censored, the ith subject contributes $Pr(y > Y_i)$ to the likelihood.

The likelihood for all n subjects is:

$$L = \prod_{i=1:\delta_i=1}^{n} f(Y_i) \prod_{i=1:\delta_i=0}^{n} S(Y_i). \tag{5.3}$$

The log-likelihood can be written as:

$$\log L = \sum_{i=1:\delta_i=1}^{n} \log h(Y_i) - \sum_{i=1:\delta_i=0}^{n} H(Y_i). \tag{5.4}$$

5.2 Fixed Right Type I

Fixed right type I censoring occurs when the experiments are run only for a fixed duration - L and the lifetimes are known for those individuals whose lifetimes are $T_i \leq L$.

More precisely, consider a population of n individuals subjected to periods of know and predetermined observation $L_1, ..., L_n$, and with lifetimes $T_1, ..., T_n$. The ith individual's lifetime is observed only if $T_i \leq L_i$.

For instance, the study stopped on a specified date, but different individuals start at different specified times.

The difference between type I and type II is that in type I censoring the number of observed lifetimes is a random variable and in type II censoring, the random is lifetimes.

Notation for type I censoring:

- n=number of individuals.
- L_i= censoring time for the ith individual.
- T_i= lifetime for the ith individual.

Not necessarily observe T_i. What observe is t_i:

$$t_i = min(T_i, L_i)$$
$$\delta_i = \begin{cases} 1 & T_i \leq L_i \\ 0 & T_i > L_i \end{cases}$$

Assumption: The $T_i s$ are *i.i.d.* with PDF $f(t)$ and survivor function *S(t)*

The *joint* pdf for t_i and δ_i is:

$$\text{Prob}\,(t_i, \delta_i) = f(t_i)^{\delta_i} S(L_i)^{1-\delta_i} \tag{5.5}$$

Explanation:

$$\begin{aligned} \text{Prob}\,(t_i, \delta_i) &= \text{Prob}\,(\delta_i = 0) \\ &= \text{Prob}\,(T_i > L_i) \\ &= S(L_i) \end{aligned}$$

For $t_i < L_i$

$$\begin{aligned} \text{Prob}\,(t_i \mid \delta_i = 1) &= \text{Prob}\,(t_i \mid T_i < L_i) \\ &= \frac{f(t_i)}{1 - S(L_i)} \end{aligned}$$

From the definition of conditional probability: $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$

Thus:

$$\text{Prob}\,(t_i = L_i, \delta_i = 0) = \text{Prob}\,(\delta_i = 0) = S(L_i) \tag{5.6}$$

and

$$\text{Prob}\,(t_i, \delta_i = 1) = \underbrace{\text{Prob}\,(t_i \mid \delta_i = 1)}_{\frac{f(t_i)}{1-\delta(L_i)}} \overbrace{\text{Prob}\,(\delta_i = 1)}^{1-\delta(L_i)} \tag{5.7}$$

$$= f(t_i) \tag{5.8}$$

combining equations 5.6 and 5.7 gives eq. 5.5

Now given n independent pairs $(t_i, \delta_i), i = 1, ..., n$, the joint pdf is:

$$f_n(t_1, \delta_1, ..., t_n, \delta_n) = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(L_i)^{1-\delta_i}$$

This is the likelihood function, $L(\lambda)$.

5.3 Random Right type I

In type I censoring its assume the censoring times $L1, \ldots, L_n$ are known and predetermined; In random censoring the individuals start at random times and both the lifetimes and the censoring times are random.

Denote by C the censoring process and by $C_1, C_2, ..., C_n$ the (potential) censoring times. The observed data are the minimum of the survival time and censoring time for each subject in the sample and the indication whether or not the subject is censored.

Statistically, the observed data $(X_i, \delta_i), i = 1, 2, ...n$, where, $X_i = min(T_i, C_i)$,:

$$\delta_i = I(T_i \leq C_i) = \begin{cases} 1 & T_i \leq C_i \quad \text{(observed failure)} \\ 0 & T_i > C_i \quad \text{(observed censoring)} \end{cases}$$

The potential data are $(T_1, C_1), (T_2, C_2), ..., (T_n, C_n)$, but the actual observed data are $(X_1, \delta_1), (X_2, \delta_2), ..., (X_n, \delta_n)$.

The interested are in making inference on the random variable T, i.e., any one of following functions

$$f(t) = \textit{density function}$$
$$F(t) = \textit{distribution function}$$
$$S(t) = \textit{survival function}$$
$$\lambda(t) = \textit{hazard function}$$

Usually, the density function *f(t)* of T may be governed by some parameters θ and *g(t)* or by some other parameters. In these case, our interested is in making inference on θ. [1]

In order to derive the density of (X, Δ), assume independent censoring, i.e., random variables T and C are independent. The density function of (X, Δ) is defined as:

$$f(x, \delta) = \lim_{h \to 0} \frac{P[x \leq X < x + h, \Delta = \delta]}{h}, \quad x \geq 0, \delta = \{0, 1\}$$

1. Case 1. $\delta = 1, i.e., T \leq C, X = min(T, C) = T$, result:

$$\begin{aligned} &P[x \leq X < x + h, \Delta = 1] \\ &= P[x \leq T < x + h, C \geq T] \\ &\approx P[x \leq T < x + h, C \geq x] \quad \text{(Note: x is a fixed number)} \\ &= P[x \leq T < x + h] * P[C \geq x] \quad \text{(by independence of T and C)} \\ &= f(\xi)h * H(x), \quad \xi \in [x, x + h), \quad \text{(Note: H(x) is the survival fucntion of C).} \end{aligned} \tag{5.9}$$

Therefore:

$$\begin{aligned} f(x, \delta = 1) &= \lim_{h \to 0} \frac{P[x \leq X < x + h, \Delta = 1]}{h} \\ &= \lim_{h \to 0} \frac{f(\xi)h * H(x)}{h} \\ &= f_T(x) H_C(x). \end{aligned} \tag{5.10}$$

2. Case 2. $\delta = 0, i.e., \quad T > C, \quad X = min(T, C) = C$, result:

$$\begin{aligned} &P[x \leq X < x + h, \Delta = 0] \\ &= P[x \leq C < x + h, T \geq C] \\ &\approx P[x \leq C < x + h, T \geq x] \quad \text{(Note: x is a fixed number)} \\ &= P[x \leq C < x + h] * P[T \geq x] \quad \text{(by independence of T and C)} \\ &= g_c(\xi)h * S(x), \quad \xi \in [x, x + h). \end{aligned} \tag{5.11}$$

[1] Notation: The density expression *f(t)* of T is different from $f(x, \delta)$ of $f(x, \Delta)$. The formal and complete expression is use $f_T(t)$ for T and $f_{X,\Delta}(X, \delta)$ for (X, Δ). But when there is no ambiguity, in the future suppress the subscripts.

Therefore

$$\begin{aligned} f(x, \delta = 0) &= \lim_{h \to 0} \frac{P[x \leq X < x + h, \Delta = 0]}{h} \\ &= \lim_{h \to 0} \frac{g_C(\xi) h * S(x)}{h} \\ &= g_C(x) S(x). \end{aligned} \tag{5.12}$$

Combining these two cases, the density function of (X, Δ):

$$\begin{aligned} f(x, \delta) &= [f_T(x) H_C(x)]^{\delta} [g_C(x) S(x)]^{1-\delta} \\ &= \{[f_T(x)]^{\delta} [S(x)]^{1-\delta}\} \{[g_C(x)]^{1-\delta} [H_C(x)]^{\delta}\} \end{aligned} \tag{5.13}$$

Sometimes it may be useful to use hazard functions. Recalling that the hazard function

$$\lambda_T(x) = \frac{f_T(x)}{S_T(x)}, \quad or \quad f_T(x) = \lambda_T(x) * S_T(x)$$

we can write $[f_T(x)]^{\delta} [S(x)]^{1-\delta}$ as

$$[f_T(x)]^{\delta} [S(x)]^{1-\delta} = [\lambda_T(x) * S_T(x)]^{\delta} * [S(x)]^{1-\delta} = [\lambda_T(x)]^{\delta} * [S(x)]$$

The likelihood function for a parametric model given our observed data (x_i, δ_i) (under independence of T and C): *i = 1, 2,..., n.*

$$L(\theta, \phi; x, \delta) = \prod_{i=1}^{n} \{[f(x_i; \theta)]^{\delta_i} [S(x_i; \theta)]^{1-\delta_i}\} \{[g(x_i; \phi)]^{1-\delta_i} [H(x_i; \theta)]^{\delta_i}$$

The mainly interested is in making inference on the parameters characterizing the distribution of T. So if θ and ϕ have no common parameters, can be used the following likelihood function to make inference on θ:

$$L(\theta; x, \delta) = \prod_{i=1}^{n} [f(x_i; \theta)]^{\delta_i} [S(x_i; \theta)]^{1-\delta_i}. \tag{5.14}$$

or equivalent,

$$L(\theta; x, \delta) = \prod_{i=1}^{n} [\lambda(x_i; \theta)]^{\delta_i} [S(x_i; \theta)] \tag{5.15}$$

Another form to express the random censor data, consider:
T_i = lifetime of the ith individual.
L_i = censoring time of ith individual.

Assume:
T_i and L_i are independent random variable.
$T1, \ldots, T_n$ are *iid* with pdf *f(t)* and survivor function *S(t)*.
$L1, \ldots, L_n$ are *iid* with pdf *g(t)* and survivor function *G(t)*.

That is:

$$\begin{aligned}
\text{Prob}\,(T) &= f(t) \\
\text{Prob}\,(T > t) &= S(t) \\
\text{Prob}\,(L) &= g(t) \\
\text{Prob}\,(L > l) &= G(t)
\end{aligned}$$

Define as before:

$$t_i = min(T_i, L_i)$$

$$\delta_i = \begin{cases} 1 & T_i \leq L_i \\ 0 & T_i > L_i \end{cases}$$

The pdf for (t_i, δ_i) is:

$$\begin{aligned}
\text{Prob}\,(t_i = t, \delta_i = 0) &= \text{Prob}\,(L_i = t, T_i > L_i) \\
&= g(s)S(t) \\
\text{Prob}\,(t_i = t, \delta_i = 1) &= \text{Prob}\,(L_i = t, T_i \leq L_i) \\
&= f(s)G(t)
\end{aligned}$$

Combining the two equations, result:

$$\text{Prob}\,(t_i = t, \delta_i) = [f(t)G(t)]^{\delta_i}[g(t)S(t)]^{1-\delta_i} \tag{5.16}$$

So, for n individuals with observations $(t_1, \delta_1), ..., (t_n \delta_n)$ the likelihood functions is:

$$\begin{aligned} L(\delta) &= \prod_{i=1}^{n} [f(t_i)G(t_i)]^{\delta_i} [g(t_i)S(t_i)]^{1-\delta_i} \\ &= \underbrace{\left(\prod_{i=1}^{n} G(t_i)^{\delta_i} g(t_i)^{1-\delta_i}\right)}_{\text{Dependes on censored r.v.s}} \underbrace{\left(\prod_{i=1}^{n} f(t_i)^{\delta_i} S(t)^{1-\delta_i}\right)}_{\text{Dependes on lifetime r.v.s}} \end{aligned}$$

It may happens that G and g, which express the censoring random variables, do not depend on parameters of interest. In that case, the likelihood function is effectively the same as the likelihood function for fixed right type I censoring.

5.4 Type II censor data

A type II censored sample is one for which:

1. Only the r smallest observation in a sample of size n are observed, $1 \ll r \ll n$;
2. r is determined **before** the data are collected.

Let the n lifetimes of the size-n sample be $T_1, ...T_n$. Their order statistics are:

$$T_{(1)} \leq T_{(2)} \leq \cdots \leq T_{(n)}$$

In type II censoring we know only the values: $T_{(1)} \dots T_{(r)}$
Let f(t) be the pdf of the lifetime:

$$f(t)dt = \text{probability of end-of-life } T \in [t, t + dt] \tag{5.17}$$

The "survivor function" or "probabilistic reliability" is:

$$\begin{aligned} S(t) &= \text{Prob}\,(T \geq t) \\ &= \int_t^{\infty} f(s)ds \end{aligned}$$

If $T_1 \dots T_n$ are *iid (independent and identically distributed)* with lifetime pdf *f(t)* and survivor function *S(t)*, then the joint pdf of $T_{(1)}, \dots, T_{(r)}$ is:

$$f_n(t_{(1)}, \dots, t_{(r)}) = \frac{n!}{(n-r)!} f(t_{(1)}) \dots f(t_{(r)}) [S(t_r)]^{n-r} \tag{5.18}$$

Explanation:

1. $\frac{n!}{(n-r)!}$ = number of ways of choosing n-r out of n items, without regard to the order in which the items are chosen.

2. For instance, n=3 and $n-r=2$: $\frac{3!}{2!}=3$. Let the items be A,B, C. The following three couples can be found: $\{A,B\};\{A,C\};\{B,C\}$.

3. $S(T_{(r)})$ = probability that a specific item will live at least $T_{(r)}$.

4. Thus $[S(t_r)]^{n-r}$ = probability that n-r specific items will have lifetimes $\geq T_{(r)}$.

5. Thus $\frac{n!}{(n-r)!}[S(t_r)]^{n-r}$= probability that n-r items, from a population of size n, will have lifetimes $\geq T_{(r)}$.

6. $f(t_{(1)})\dots f(t_{(r)})$ = the joint probability density for the r specific independent items whose lifetimes are known.

Example 5.4.1:
Exponential distribution

Suppose, as before that t is exponentially distributed:

$$f(t)=\lambda e^{-\lambda t}, \quad t\geq 0 \tag{5.19}$$
$$S(t)=e^{-\lambda t} \tag{5.20}$$

The likelihood function becomes:

$$L(\lambda)=\prod_{i=1}^{n}(\lambda e^{-\lambda t})^{\delta_i}e^{-\lambda t(1-\delta_i)} \tag{5.21}$$
$$=\lambda^r \exp\left(-\sum_{i=1}^{n} t_i\right) \tag{5.22}$$
$$\tag{5.23}$$

where $r=\sum_{i=1}^{n}\delta_i$ is number of observed "deaths" or failures. What is the MLE of λ? Let $T=\sum_{i=1}^{n} t_i$, so $L(\lambda)=\lambda^r e^{-\lambda t}$. Thus:

$$0=\frac{dL}{d\lambda}=e^{-\lambda t}[re^{r-1}-\lambda^r T]\Rightarrow \hat{\lambda}=\frac{r}{T}$$

Compare the likelihood functions for types I and II censoring:

$$L_{II} = \frac{n!}{(n-r)!} f(t_{(1)}) \dots f(t_{(r)})[S(t_r)]^{n-r}$$
$$L_I = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(L_i)^{1-\delta_i}$$

For L_1:
Each observed lifetime ($\delta_i = 1$) contributed a factor $f(t_i)$.
Each observed lifetime ($\delta_i = 0$) contributed a factor$S(L_i)$.
Thus L_I is similar in form to L_{II}, though different in origin and precise structure.

5.5 Complete data MLE Weibull

For complete data and Weibull distribution, the scale and shape parameters are η and β respectively, given the followed equation:

$$L(\eta, \beta) = \left(\frac{\beta}{\eta^\beta}\right)^n \prod_{i=1}^{n} \left[t_i^{\beta-1} \exp\left(-\left(\frac{t_i}{\eta}\right)^\beta\right)\right] \tag{5.24}$$

The log-likelihood is obtained by:

$$\begin{aligned}
\ln L(\eta, \beta) &= \sum_{i=1}^{n} \left\{ \ln \left[\frac{\beta}{\eta} \left(\frac{t_i}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{t_i}{\eta}\right)^\beta\right)\right]\right\} \\
&= \sum_{i=1}^{n} \left\{ \ln\left(\frac{\beta}{\eta}\right) + (\beta-1)\ln\left(\frac{t_i}{\eta}\right) + \left(-\frac{t_i}{\eta}\right)^\beta \right\} \\
&= \sum_{i=1}^{n} \left\{ \ln\beta - \ln\eta + (\beta-1)\ln t_i - (\beta-1)\ln\eta - \left(\frac{t_i}{\eta}\right)^\beta \right\} \\
&= \sum_{i=1}^{n} \left\{ \ln\beta - \ln\eta + (\beta-1)\ln t_i - \beta\ln\eta + \ln\eta - \left(\frac{t_i}{\eta}\right)^\beta \right\} \\
&= \sum_{i=1}^{n} \left\{ \ln\beta - \beta\ln\eta + (\beta-1)\ln t_i - \left(\frac{t_i}{\eta}\right)^\beta \right\}
\end{aligned}$$

and result finally after some simplification:

$$\ln L(\eta, \beta) = n\ln\beta - n\beta\ln\eta + (\beta-1)\sum_{i=1}^{n}\ln t_i - \sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^\beta \tag{5.25}$$

Differentiating 5.25 with respect to η we have:

$$\begin{aligned}
\frac{\partial l(\eta,\beta)}{\partial \eta} &= 0 - n\beta(\frac{1}{\eta}) + 0 - \beta\sum_{i=1}^{n}\left\{\left(\frac{t_i}{\eta}\right)^{\beta-1}\left(-\frac{t_i}{\eta^2}\right)\right\} \\
&= -\frac{n\beta}{\eta} + \beta\sum_{i=1}^{n}\left(\frac{t_i^{\beta-1}t_i}{\eta^{\beta-1+2}}\right) \\
&= -\frac{n\beta}{\eta} + \beta\sum_{i=1}^{n}\left(\frac{t_i^{\beta-1+1}}{\eta^{\beta+1}}\right) \\
&= -\frac{n\beta}{\eta} + \frac{\beta}{\eta}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} \qquad (5.26)
\end{aligned}$$

Differentiating 5.46 with respect to β we have:

$$\frac{\partial l(\eta,\beta)}{\partial \beta} = \frac{n}{\beta} - n\ln\eta + \sum_{i=1}^{n}\ln t_i - \sum_{i=1}^{n}\left\{\left(\frac{t_i}{\eta}\right)^{\beta}\ln\left(\frac{t_i}{\eta}\right)\right\} \qquad (5.27)$$

The second derivative in order to η

$$\begin{aligned}
\frac{\partial^2 l(\eta,\beta)}{\partial \eta^2} &= \frac{n\beta}{\eta^2} - \frac{\beta}{\eta^2}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} + \frac{\beta}{\eta}\beta\sum_{i=1}^{n}\left\{\left(\frac{t_i}{\eta}\right)^{(\beta-1)}\left(-\frac{t_i}{\eta^2}\right)\right\} \\
&= \frac{n\beta}{\eta^2} - \frac{\beta}{\eta^2}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} - \frac{\beta}{\eta^2}\beta\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{(\beta)} \\
&= \frac{n\beta}{\eta^2} - \frac{\beta}{\eta^2}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta}(1+\beta) \\
&= \frac{\beta}{\eta^2}\left[n - (\beta+1)\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta}\right]
\end{aligned}$$

and the second derivative in order to β

$$\begin{aligned}
\frac{\partial^2 l(\eta,\beta)}{\partial \beta^2} &= -\frac{n}{\beta^2} + 0 + 0 - \sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{\beta}\ln\left(\frac{t_i}{\eta}\right)\ln\left(\frac{t_i}{\eta}\right)\right] \\
&= -\frac{n}{\beta^2} - \sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{\beta}\left(\ln\left(\frac{t_i}{\eta}\right)\right)^2\right]
\end{aligned}$$

finally differentiation both η and β, we have:

$$\begin{aligned}
\frac{\partial^2 l(\eta,\beta)}{\partial\eta\partial\beta} &= 0 - \frac{n}{\eta} + 0 - \sum_{i=1}^{n}\left[\beta\left(\frac{t_i}{\eta}\right)^{(\beta-1)}\left(-\frac{t_i}{\eta^2}\right)\ln\left(\frac{t_i}{\eta}\right) + \left(\frac{t_i}{\eta}\right)^{\beta}\frac{\left(-\frac{t_i}{\eta^2}\right)}{\left(\frac{t_i}{\eta}\right)}\right] \\
&= -\frac{n}{\eta} + \frac{\beta}{\eta}\sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{(\beta)}\ln\left(\frac{t_i}{\eta}\right)\right] + \frac{1}{\eta}\sum_{1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} \\
&= -\frac{1}{\eta}\left[n - \beta\sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{(\beta)}\ln\left(\frac{t_i}{\eta}\right)\right] - \sum_{1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta}\right]
\end{aligned}$$

now equalizing to zero the expression 5.26:

$$\begin{aligned}
-\frac{n\beta}{\eta} + \frac{\beta}{\eta}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} &= 0 \\
\frac{1}{\eta}\left[-n\beta + \beta\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta}\right] &= 0 \\
-n\beta + \beta\frac{1}{\eta^{\beta}}\sum_{i=1}^{n}{t_i}^{\beta} &= 0
\end{aligned}$$

and simplify:

$$\begin{aligned}
\eta^{\beta} &= -\frac{\beta\sum_{i=1}^{n}t_i^{\beta}}{n\beta} \qquad (5.28) \\
\eta &= \sqrt[\beta]{\frac{\sum_{i=1}^{n}t_i^{\beta}}{n}} \\
\eta &= \left(\frac{1}{n}\sum_{i=1}^{n}{t_i}^{\beta}\right)^{\frac{1}{\beta}}
\end{aligned}$$

Substituting 5.29 into 5.27 we have:

$$
\begin{aligned}
\frac{n}{\beta} - n\ln\eta + \sum_{i=1}^{n}\ln t_i - \sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{\beta}(\ln t_i - \ln\eta)\right] &= 0 \\
\frac{n}{\beta} - n\ln\eta + \sum_{i=1}^{n}\ln t_i - \sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln t_i + \sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln\eta &= 0 \\
\frac{1}{\beta} - \ln\eta + \frac{1}{n}\sum_{i=1}^{n}\ln t_i - \frac{1}{n}\sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}} + \frac{1}{n}\sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln\eta &= 0 \quad \textit{and} \quad \frac{1}{n}\sum_{i=1}^{n}t_i^{\beta} = \eta^{\beta} \\
\frac{1}{\beta} - \ln\eta + \frac{1}{n}\sum_{i=1}^{n}\ln t_i - \frac{1}{n}\sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln t_i + \frac{\eta^{\beta}}{\eta^{\beta}}\ln\eta &= 0 \\
\frac{1}{\beta} - \ln\eta + \frac{1}{n}\sum_{i=1}^{n}\ln t_i - \frac{1}{n}\frac{\sum_{i=1}^{n}t_i^{\beta}\ln t_i}{\frac{\sum_{i=1}^{n}\ln t_i^{\beta}}{n}} + \ln\eta &= 0 \\
\frac{1}{\beta} + \frac{1}{n}\sum_{i=1}^{n}\ln t_i - \frac{\sum_{i=1}^{n}t_i^{\beta}\ln t_i}{\sum_{i=1}^{n}t_i^{\beta}} &= 0
\end{aligned}
\tag{5.29}
$$

5.5.1 Simulation complete data MLE Weibull

To compute a simulation of *MLE* for data complete using Weibull distribution the following algorithm is developed:

Step 1: Define the parameters of distribution and initialize the variable to use

Step 2: Define the inverse function in order to Uniform random number U_i that give our random time t_i

Step 3: Generate U_i random uniform (0,1)

Step 4: Generate t_i from step (1) and step (4), or directly from random distribution function and repeat for n times (the dimension of the sample)

Step 5: Calculate the MLE using the equation 5.27 and 5.39 and a solver optimization algorithm *(nsolver from numpy)*

Step 6: Repeat step (3) to step (5)

Step 7: Compare each t_i with t_c and calculate the *erf*

And the program has been written in python:

```
for i in range(0,len(beta)):
    sh=beta[i]
    solsh=[]; solsc=[]
    for f in range(0,m):
        Xi = sc*np.random.weibull(sh,n)
        Xilog=np.log(Xi)
        Xis=np.sum(Xilog)
        lti1=(x/r)**b
        f1=-n*b/r+b/r*sum([lti1.subs(x,i) for i in Xi])
        lti2=(x/r)**b*sp.log(x/r)
        f2=n/b-n*sp.log(r)+Xis- sum([lti2.subs(x,i) for i in Xi])
        rr=beta[i]-0.4
        res=nsolve((f1,f2), (b,r), (rr, 7),solver='bisect',
                        verify=False)
        solsh.append(res[0])
        solsc.append(res[1])
```

Algorithm 5: MLE for complete data with python (partial)

β	μ_β	σ	ξ	μ_η	σ	ξ
$\beta_{0.5}$	0.50	0.08	0	10.60	4.36	0.1
β_1	1.05	0.14	0.05	10.00	1.65	0
$\beta_{1.5}$	1.69	0.36	0.13	10.10	0.90	0
β_2	2.01	0.36	0	9.91	1.16	0

Table 5.1: Simulation Complete Data, Weibull, (β), $\eta = 10$, $n = 100$

The simulation was made for different shape factor β - *0.5 to 2* and the scale parameter η is equal to *10*; The number of sample is $n = 100$ and the number of cycle simulation is $M = 1000$; some of the results are show in table 5.1. The algorithm gives a good precision in estimate the shape factor β and low standard deviation; the largest deviation corresponds to the $\beta = 1.5$. To the estimation of the scale factor the standard deviation is greater than shape factor, but, the percentage error is very small. To final conclusion the algorithm is precise, faster and easy to program. This algorithm validate the routines, functions and packages *(numpy)* used in python language program.

5.6 Right censored type I MLE Weibull

Suppose that $t_1 \leq ... \leq t_r$ is know to have failed during the study and the remaining $t_n - t_r = t_q$ censored but the censored units do not have all the same censoring time, then the likelihood function of the 2-parameter Weibull distribution is:

$$L(\eta, \beta) = \prod_{i=1}^{n} \left\{ \frac{\beta}{\eta} \left(\frac{t_i}{\eta} \right)^{\beta-1} \exp\left(- \left(\frac{t_i}{\eta} \right)^{\beta} \right) \right\}^{\delta_i} \left\{ \exp\left(\left(-\frac{t_i}{\eta} \right)^{\beta} \right) \right\}^{1-\delta_i} \tag{5.30}$$

And the log-likelihood is:

$$\begin{aligned}
\ln L(\eta, \beta) &= \sum_{i=1}^{n} \left\{ \ln \left[\frac{\beta}{\eta} \left(\frac{t_i}{\eta} \right)^{\beta-1} \exp\left(- \left(\frac{t_i}{\eta} \right)^{\beta} \right) \right]^{\delta_i} \right\} + \sum_{i=1}^{n} \left\{ \ln \left[\exp\left(-\frac{t_i}{\eta} \right)^{\beta} \right]^{1-\delta_i} \right\} \\
&= \sum_{i=1}^{n} \left\{ \delta_i \left[\ln\left(\frac{\beta}{\eta} \right) + (\beta - 1) \ln\left(\frac{t_i}{\eta} \right) + \left(-\frac{t_i}{\eta} \right)^{\beta} \right] \right\} + \sum_{i=1}^{n} \left\{ (1 - \delta_i) \left[\left(-\frac{t_i}{\eta} \right)^{\beta} \right] \right\} \\
&= \sum_{i=1}^{n} \left\{ \delta_i \left[\ln \beta - \ln \eta + (\beta - 1) \ln t_i - (\beta - 1) \ln \eta - \left(\frac{t_i}{\eta} \right)^{\beta} \right] \right\} + \sum_{i=1}^{n} \left\{ (1 - \delta_i) \left[\left(-\frac{t_i}{\eta} \right)^{\beta} \right] \right\} \\
&= \sum_{i=1}^{n} \left\{ \delta_i \left[\ln \beta - \ln \eta + \beta \ln t_i - \ln t_i - \beta \ln \eta + \ln \eta - \left(\frac{t_i}{\eta} \right)^{\beta} \right] - \left(\frac{t_i}{\eta} \right)^{\beta} + \delta_i \left(\frac{t_i}{\eta} \right)^{\beta} \right\} \\
&= \sum_{i=1}^{n} \left\{ \delta_i \ln \beta + \delta_i \beta \ln t_i - \delta_i \ln t_i - \delta_i \beta \ln \eta - \delta_i \left(\frac{t_i}{\eta} \right)^{\beta} - \left(\frac{t_i}{\eta} \right)^{\beta} + \delta_i \left(\frac{t_i}{\eta} \right)^{\beta} \right\}
\end{aligned}$$

and

$$\sum_{i=1}^{n} \delta_i = r$$

which became:

$$\ln L(\eta, \beta) = r \ln \beta - r\beta \ln \eta + (\beta - 1) \sum_{i=1}^{n} (\delta_i \ln t_i) - \sum_{i=1}^{n} \left(\frac{t_i}{\eta} \right)^{\beta} \tag{5.31}$$

Differentiating 5.31 with respect to η give the following:

$$\begin{aligned}\frac{\partial l(\eta,\beta)}{\partial \eta} &= 0 - \frac{r\beta}{\eta} + 0 - \beta \sum_{i=1}^{n} \left[\left(\frac{t_i}{\eta}\right)^{\beta-1} \left(-\frac{t_i}{\eta^2}\right) \right] \\ &= -\frac{r\beta}{\eta} + \beta \sum_{i=1}^{n} \left(\frac{{t_i}^{\beta}}{\eta^{\beta+1}} \right) \\ &= -\frac{r\beta}{\eta} + \frac{\beta}{\eta^{\beta+1}} \sum_{i=1}^{n} (t_i)^{\beta}\end{aligned} \tag{5.32}$$

Differentiating 5.31 with respect to β give the following:

$$\frac{\partial l(\eta,\beta)}{\partial \beta} = \frac{r}{\beta} - r\ln\eta + \sum_{i=1}^{n}(\delta_i \ln t_i) - \sum_{i=1}^{n} \left[\left(\frac{t_i}{\eta}\right)^{\beta} \ln\left(\frac{t_i}{\eta}\right) \right] \tag{5.33}$$

it's yields:

$$\frac{\partial l(\eta,\beta)}{\partial \beta} = \frac{r}{\beta} - r\ln\eta + r\sum_{i=1}^{n}(\ln t_i) - \frac{1}{\eta^{\beta}} \sum_{i=1}^{n} t_i \ln t_i + \frac{\ln\eta}{\eta^{\beta}} \sum_{i=1}^{n} t_i \ln t_i \tag{5.34}$$

The second derivative in order to η

$$\begin{aligned}\frac{\partial^2 l(\eta,\beta)}{\partial \eta^2} &= \frac{r\beta}{\eta^2} - \frac{\beta}{\eta^2} \sum_{i=1}^{n} \left(\frac{t_i}{\eta}\right)^{\beta} + \frac{\beta}{\eta}\beta \sum_{i=1}^{n} \left[\left(\frac{t_i}{\eta}\right)^{\beta-1} \left(-\frac{t_i}{\eta^2}\right) \right] \\ &= \frac{r\beta}{\eta^2} - \frac{\beta}{\eta^2} \sum_{i=1}^{n} \left(\frac{t_i}{\eta}\right)^{\beta} (1+\beta) \\ &= \frac{\beta}{\eta^2} \left[r - (\beta+1) \sum_{i=1}^{n} \left(\frac{t_i}{\eta}\right)^{\beta} \right]\end{aligned}$$

The second derivative in order to β

$$\begin{aligned}\frac{\partial^2 l(\eta,\beta)}{\partial \beta^2} &= -\frac{r}{\beta^2} + 0 + 0 - \sum_{i=1}^{n} \left[\left(\frac{t_i}{\eta}\right)^{\beta} \ln\left(\frac{t_i}{\eta}\right) \ln\left(\frac{t_i}{\eta}\right) \right] \\ &= -\frac{r}{\beta^2} - \sum_{i=1}^{n} \left[\left(\frac{t_i}{\eta}\right)^{\beta} \left(\ln\left(\frac{t_i}{\eta}\right) \right)^2 \right]\end{aligned}$$

finally differentiating both η and β results:

$$\begin{aligned}
\frac{\partial^2 l(\eta,\beta)}{\partial\eta\partial\beta} &= 0 - \frac{r}{\eta} + 0 - \sum_{i=1}^{n}\left[\beta\left(\frac{t_i}{\eta}\right)^{\beta-1}\ln\left(-\frac{t_i}{\eta^2}\right)\ln\left(\frac{t_i}{\eta}\right) + \left(\frac{t_i}{\eta}\right)^{\beta}\frac{-\frac{t_i}{\eta^2}}{\frac{t_i}{\eta}}\right] \\
&= -\frac{r}{\eta} + \frac{\beta}{\eta}\sum_{i=1}^{n}\left[\beta\left(\frac{t_i}{\eta}\right)^{\beta}\ln\left(\frac{t_i}{\eta}\right)\right] + \frac{1}{\eta}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} \\
&= -\frac{1}{\eta}\left[r - \beta\sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{\beta}\ln\left(\frac{t_i}{\eta}\right)\right] - \sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta}\right]
\end{aligned}$$

equating 5.34 to zero

$$\begin{aligned}
-\frac{n\delta_i\beta}{\eta} + \frac{\beta}{\eta}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} &= 0 \\
-r\beta + \beta\frac{1}{\eta^{\beta}}\sum_{i=1}^{n}(t_i)^{\beta} &= 0 \\
-r\beta\eta^{\beta} + \beta\sum_{i=1}^{n}(t_i)^{\beta} &= 0 \\
\eta^{\beta} &= \frac{\beta\sum_{i=1}^{n}(t_i)^{\beta}}{r\beta} \\
\eta &= \sqrt[\beta]{\frac{1}{r}\sum_{i=1}^{n}(t_i)^{\beta}} \\
\eta &= \left(\frac{1}{r}\sum_{i=1}^{n}(t_i)^{\beta}\right)^{\frac{1}{\beta}} \qquad (5.35)
\end{aligned}$$

Substituting 5.35 into 5.33 we have:

$$
\begin{aligned}
\frac{r}{\beta} + r\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - \sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta}[\ln t_i - \ln\eta] &= 0\\
\frac{r}{\beta} + r\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - \sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln t_i + \sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln\eta &= 0\\
\frac{r}{\beta} + r\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - \sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln t_i + \sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln\eta &= 0\\
\frac{r}{\beta} + r\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - \sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln t_i + \sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln\eta &= 0\\
\frac{r}{\beta} + r\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - \frac{\sum_{i=1}^{n} t_i^{\beta}}{\frac{\sum_{i=1}^{n} t_i^{\beta}}{\delta_i n}} + \ln\eta &= 0\\
\frac{r}{\beta} + (1-r)\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - \frac{\sum_{i=1}^{n} t_i^{\beta}\ln t_i}{\sum_{i=1}^{n} t_i^{\beta}} &= 0
\end{aligned}
\tag{5.36}
$$

5.6.1 Simulation right censored type I MLE Weibull

To compute a simulation of MLE for right censored type I MLE Weibull distribution we will use the following algorithm,

Step 1: Define the parameters of distribution and initialize the variable to use;

Step 2: Define the inverse function in order to Uniform random number U_i that give our random time t_i;

Step 3: Generate U_i random uniform (0,1);

Step 4: Generate t_i from step (1) and step (4), or directly from random distribution function and repeat for n times (the dimension of the sample);

Step 5: Calculate the MLE using the equation 5.27 and 5.39 and a solver optimization algorithm *(nsolver from numpy)*;

Step 6: Repeat step (3) to step (5);

Step 7: Compare each t_i with t_c and calculate the erf.

The program has been written in python as can see in the algorithm 6.

```
for i in range(0,len(beta)):
    sh=beta[i]
    solsh=[]; solsc=[]
    for p in range(0,m):
        #generation data
        Xi = sc*np.random.weibull(sh,n) # generate the random number
        T_cens=[None]*n;cr1=0;
        # time theoretical censor from weibull
        censoring_time = sc*(-np.log(censnumber))**(1/sh)
        #simulation times
        for j in range(0,n):
            T_cens[j] = min(censoring_time, Xi[j])
            #Count the number of complete data - r
        for j in range(0,n):
            if T_cens[j] < censoring_time:
                cr1 = cr1+1
            else:
                cr1=cr1
        #Vector with only complete data
       r=cr1
       Si= sorted(T_cens)
       Silog=np.log(Si)
       Si_sum=sum(Silog)
       #resolution of numerical equation
       lsum=x**b
       lsumlog=Si_sum
       f1=-r*b/e+b/(e**(b+1))*sum([lsum.subs(x,i) for i in Si])
       #define the members of eauation - shape
       lti2=sympy.log(x)
       f2=r/b-r*sympy.log(e)+r*lsumlog-1/(e**b)*sum([lsum.subs(x,i)
        for i in Si])*sum([lti2.subs(x,i) for i in Si])
        +sympy.log(e)/(e**b)*sum([lsum.subs(x,i) for i in Si])
       #Resolve numerical equations f1 e f2
       rr=sh-0.3
       res=nsolve((f1,f2), (b,e), (rr, 8),solver='bisect',
                        verify=False) #
       solsh.append(res[0])
       solsc.append(res[1])
```

Algorithm 6: MLE for censored data right type I and II program with python (partial)

The simulation analysis made two type of simulations:

- the first simulation is the estimate by *MLE* of the parameter shape factor β with change of the shape factor β and percentage of censored data $C\%$. The scale factor have value $\eta = 1$. The results gives average from the number of simulation cycle $M = 1000$, the standard deviation and the relative error from the original shape factor.

- The second simulations is the estimate by *MLE* of the scale factor with value $\eta = 10$ with change of the shape factor β and percentage of censored data $\%C$. The results give average from the number of simulation cycle $M = 1000$, the standard deviation and the relative error from the original shape factor.

	$\mathbf{C}_{5\%}$			$\mathbf{C}_{10\%}$			$\mathbf{C}_{20\%}$			$\mathbf{C}_{30\%}$		
β	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\beta_{0.5}$	1.5	0.8	2.1	1.5	1	2	0.96	0.78	0.92	1.2	1.1	1.5
β_1	1.2	0.84	0.18	0.88	0.62	0.12	0.64	0.46	0.36	0.44	0.33	0.56
$\beta_{1.5}$	1.1	0.62	0.29	0.96	0.7	0.36	0.38	0.25	0.75	0.77	0.41	0.49
β_2	1.4	0.86	0.29	1.1	0.79	0.47	1	0.59	0.49	1.1	0.53	0.43

Table 5.2: Simulation Right Type I, Weibull $(\beta, \%C)$, $\eta = 1$, $n = 100$

In the simulation of shape factor can see very well that when $C\%$ percentage of censored data increase the precision decrease. The percentage error is great when $\beta = 0.5$ and reduced when shape factor growth up.

	$\mathbf{C}_{5\%}$			$\mathbf{C}_{10\%}$			$\mathbf{C}_{20\%}$			$\mathbf{C}_{30\%}$		
β	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ	μ	σ	ξ
$\beta_{0.5}$	610	2,900	60	380	1,200	37	14	7.2	0.35	890	4,300	88
β_1	11	3.2	0.06	10	2.8	0.01	10	2.5	0.02	13	7.5	0.31
$\beta_{1.5}$	9	2	0.1	9.3	1.4	0.07	11	3.8	0.12	11	1.9	0.06
β_2	9.5	1.3	0.05	9.5	0.96	0.05	11	1.8	0.06	11	1.6	0.06

Table 5.3: Simulation Right Type I, Weibull $(\beta, \%C)$, $\eta = 10$, $n = 100$

In simulation of the scale factor the greatest deviation is when $\beta = 0.5$ and with the high numbers. The algorithm to shape factor $\beta = 0.5$ and scale factor $\eta = 10$ not work or fit well, but to the other values of scale factor β the simulation works fine and the results are very good.

To final conclusion the algorithm is precise, faster and easy to program (the exception is the value $\beta = 0.5$). This algorithm validate the routines, functions and packages *(numpy)* used in python language program.

5.7 Left censored MLE Weibull

Left censoring is when the event of interest has already occurred before the study or analysis. This is very rarely encountered. A left censoring scheme is, such as, the random variable of interest, X, is observed if it is greater than, or equal, to a left censoring variable L, otherwise, L is observed.

The analysis is then based on the pair of random variables (U, δ) where $U = max(L, X)$ and $\delta = 1 \quad \{L \leq X\}$. The problem concerns the estimation of the survival function $S_X(t) = Pr\{X > t\}$ from a left censored sample where X is assumed to be independent of L then the likelihood function of the two parameter Weibull distribution is:

$$L(\eta, \beta) = \prod_{i=1}^{n} \left\{ \frac{\beta}{\eta} \left(\frac{t_i}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right) \right\}^{\delta_i} \left\{ 1 - \exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right) \right\}^{1-\delta_i} \tag{5.37}$$

And the log-likelihood is:

$$\begin{aligned}
\ln L(\eta, \beta) &= \sum_{i=1}^{n} \left\{ \ln \left[\frac{\beta}{\eta} \left(\frac{t_i}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right) \right] \right\}^{\delta_i} + \sum_{i=1}^{n} \left\{ \ln \left[1 - \exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right) \right] \right\}^{1-\delta_i} \\
&= \sum_{i=1}^{n} \left\{ \delta_i \left[\ln\left(\frac{\beta}{\eta}\right) + (\beta - 1) \ln \frac{t_i}{\eta} - \left(\frac{t_i}{\eta}\right)^{\beta} \right] \right\} + \sum_{i=1}^{n} \left\{ (1 - \delta_i) \left[\ln 1 - \ln . \exp\left(-\left(\frac{t_i}{\eta}\right)^{\beta}\right) \right] \right\} \\
&= \sum_{i=1}^{n} \left\{ \delta_i \left[\ln \beta - \ln \eta + (\beta - 1) \ln t_i - (\beta - 1) \ln \eta - \left(\frac{t_i}{\eta}\right)^{\beta} \right] \right\} + \sum_{i=1}^{n} \left\{ (1 - \delta_i) \left(\frac{t_i}{\eta}\right)^{\beta} \right\} \\
&= \sum_{i=1}^{n} \left\{ \delta_i \left[\ln \beta - \ln \eta + \beta \ln t_i - \ln t_i - \beta \ln \eta + \ln \eta - \left(\frac{t_i}{\eta}\right)^{\beta} \right] + \left(\frac{t_i}{\eta}\right)^{\beta} - \delta_i \left(\frac{t_i}{\eta}\right)^{\beta} \right\} \\
&= \sum_{i=1}^{n} \left\{ \delta_i \ln \beta + \delta_i \beta \ln t_i - \delta_i \ln t_i - \delta_i \beta \ln \eta - \delta_i \left(\frac{t_i}{\eta}\right)^{\beta} + \left(\frac{t_i}{\eta}\right)^{\beta} - \delta_i \left(\frac{t_i}{\eta}\right)^{\beta} \right\}
\end{aligned}$$

which becomes:

$$\ln L(\eta, \beta) = r \ln \beta - r\beta \ln \eta + (\beta - 1) \sum_{i=1}^{n} (\delta_i \ln t_i) - (2r - 1) \sum_{i=1}^{n} \left(\frac{t_i}{\eta}\right)^{\beta} \tag{5.38}$$

Differentiating 5.38 with respect to η give the following:

$$\begin{aligned}\frac{\partial l(\eta,\beta)}{\partial \eta} &= 0 - \frac{r\beta}{\eta} + 0 - (2r-1)\beta \sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{\beta-1}\left(-\frac{t_i}{\eta^2}\right)\right] \\ &= -\frac{r\beta}{\eta} + (2r-1)\beta \sum_{i=1}^{n}\left(\frac{{t_i}^{\beta}}{\eta^{\beta+1}}\right) \\ &= -\frac{r\beta}{\eta} + (2r-1)\frac{\beta}{\eta}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} \end{aligned} \tag{5.39}$$

Differentiating 5.38 with respect to β give the following:

$$\frac{\partial l(\eta,\beta)}{\partial \beta} = \frac{r}{\beta} - r\ln\eta + \sum_{i=1}^{n}(\delta_i \ln t_i) - (2r-1)\sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{\beta}\ln\left(\frac{t_i}{\eta}\right)\right] \tag{5.40}$$

equating 5.39 to zero

$$\begin{aligned} -\frac{r\beta}{\eta} + (2r-1)\frac{\beta}{\eta}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} &= 0 \\ r\beta + (2r-1)\beta\frac{1}{\eta^{\beta}}\sum_{i=1}^{n}(t_i)^{\beta} &= 0 \\ -r\beta\eta^{\beta} + (2r-1)\beta\sum_{i=1}^{n}(t_i)^{\beta} &= 0 \\ \eta^{\beta} &= \frac{(2r-1)\beta\sum_{i=1}^{n}(t_i)^{\beta}}{r\beta} \\ \eta &= \sqrt[\beta]{\frac{(2r-1)}{r}\sum_{i=1}^{n}(t_i)^{\beta}} \\ \eta &= \left(\frac{(2r-1)}{r}\sum_{i=1}^{n}(t_i)^{\beta}\right)^{\frac{1}{\beta}} \end{aligned} \tag{5.41}$$

Substituting 5.41 into 5.40 we have:

$$\frac{r}{\beta} - r\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - (2r-1)\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta}[\ln t_i - \ln\eta] = 0$$

$$\frac{r}{\beta} - r\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - (2r-1)\sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln t_i + (2r-1)\sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln\eta = 0$$

$$\frac{r}{\beta} - r\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - (2r-1)\sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln t_i + (2r-1)\sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln\eta = 0 \quad \textit{Substituting} \quad \eta^{\beta}$$

$$\frac{r}{\beta} - r\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - (2r_i-1)\sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln t_i + (2r-1)\sum_{i=1}^{n}\frac{t_i^{\beta}}{\eta^{\beta}}\ln\eta = 0$$

$$\frac{r}{\beta} - r\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - (2r-1)\frac{\sum_{i=1}^{n}t_i^{\beta}}{\frac{\sum_{i=1}^{n}t_i^{\beta}}{r}} + (2r-1)\ln\eta = 0$$

$$\frac{r}{\beta} + (3r-1)\ln\eta + \sum_{i=1}^{n}\delta_i \ln t_i - (2r-1)\frac{\sum_{i=1}^{n}t_i^{\beta}\ln t_i}{\sum_{i=1}^{n}t_i^{\beta}} = 0$$

The second derivative in order to η

$$\begin{aligned}
\frac{\partial^2 l(\eta,\beta)}{\partial\eta^2} &= \frac{r\beta}{\eta^2} - (2r-1)\frac{\beta}{\eta^2}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} + (2r-1)\frac{\beta}{\eta}\beta\sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{\beta-1}\left(-\frac{t_i}{\eta^2}\right)\right] \\
&= \frac{r\beta}{\eta^2} - \frac{\beta}{\eta^2}\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta}(1+\beta)(2r-1) \\
&= \frac{\beta}{\eta^2}\left[r - (2r-1)(\beta+1)\sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta}\right]
\end{aligned}$$

The second derivative in order to β

$$\begin{aligned}
\frac{\partial^2 l(\eta,\beta)}{\partial\beta^2} &= -\frac{r}{\beta^2} + 0 + 0 - (2r-1)\sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{\beta}\ln\left(\frac{t_i}{\eta}\right)\ln\left(\frac{t_i}{\eta}\right)\right] \\
&= -\frac{r}{\beta^2} - (2r-1)\sum_{i=1}^{n}\left[\left(\frac{t_i}{\eta}\right)^{\beta}\ln\left(\frac{t_i}{\eta}\right)^2\right]
\end{aligned}$$

5.8 The estimation parametric method Expectation Maximization - EM

The Expectation Maximization method (EM) is an iterative process that can be used to calculate the maximum likelihood estimators when the study have incomplete data or censored data. The designation was given by *Dempster, Laird and Rubin* in Dempster et al. (1977), although the underlying idea was outlined in 1972 by *Orchard and Woodbury.*

The algorithm consists of two steps performed repeatedly until a convergence criteria is met. The EM algorithm is used in a wide range of statistical applications because its formulation, which reduces the complexity of the estimation problem. One of the applications of the algorithm is when the maximum likelihood estimator has to be calculated in the presence of incomplete data. The aim of the algorithm is basically to simplify an incomplete data problem to a complete data problem, which is often easier to solve, creating a connection between the two conditions.

The EM algorithm is a process that converges to the maximum likelihood estimator and is based on the substitution of a difficult maximization of likelihood by a sequence of easier maximizations whose limit is the response to the original problem.

The likelihood function for complete data is usually less complex compared to the function for incomplete data, Balakrishnan and Mitra (2012).

One of the negative aspects of the EM algorithm is its slow convergence. However, other algorithms based on the EM algorithm have been proposed to increase the speed of convergence and preserve its simplicity, namely the Incremental EM algorithm (EMI) and the Sparse EM algorithm (SPEM).

Let X be the set of data complete with the probability density function $f_c(x, \theta)$ and θ the parameters that characterize the distribution. The log-likelihood function corresponding to the complete sample is represented by,

$$\ln L_c(x, \theta) = l_c(x, \theta) \tag{5.42}$$

In the presence of incomplete data some events are not observed. Let Y be the set of observed data and Z be the set of unknown data, X can be represented as a function of (y, z), so that:

Each iteration of the EM algorithm involves two steps, step E (expectation) and step M (maximization), defined by *Mclachlan (2008)*,

Step E: Calculate $Q(\theta \mid \theta^{(K)})$ where,

$$Q(\theta \mid \theta^{(K)}) = E_{(K)}\left[l_c(x,\theta), \delta\theta^{(K-1)}\right] \tag{5.43}$$

Step M: Finding $\theta^{(K+1)}$that maximizes $Q(\theta \mid \theta^{(K)})$ this is,

$$\theta^{(K+1)=argmaxQ(\theta|\theta^{(K)})}(5.44)$$

$$Q(\theta^{(K+1)}, \theta^{(K)}) \geq Q(\theta, \theta^{(K)}) \tag{5.45}$$

The procedure is performed until the difference between iteration k and iteration k + 1, decreases to an acceptable value, $\epsilon > 0$.

Step E of the algorithm calculates the conditional expected value of the *log* of the likelihood function for complete data given the observed sample and step *M* finds its maximum.

This algorithm requires an initial solution for the values of the distribution parameters, called $\theta^{(0)}$. The choice of this initial solution requires particular attention as the speed of convergence of the algorithm can become extremely slow due to poor choice. Another aspect to consider is that the maximum likelihood equation can have multiple solutions corresponding to local maxima, so the choice of the initial solution becomes important.

5.8.1 The method EM to right censored

The data vector, Z_i, can be considered as the not observed data. The observed data (Y_i, δ_i) in $Y_i = min(T_i, C_i)$, where C_i corresponds to observation time, $\delta_i = 1$ where $(T_i \leq C_i)$ and uncensored data correspond $\delta_i = 0$ where $(T_i > Y_i)$.

Step *E* of the algorithm requires the calculation of the conditional expected value of log-likelihood for complete data given the observed sample. In this case the log-likelihood function for complete data and the logarithm equation to the MLE for complete is given by:

$$\ln L(\eta, \beta) = n \ln \beta - n\beta \ln \eta + (\beta - 1)\sum_{i=1}^{n}(\ln t_i) - \sum_{i=1}^{n}\left(\frac{t_i}{\eta}\right)^{\beta} \tag{5.46}$$

The distribution parameters, $\delta_i = (\delta_1, \delta_2, ..., \delta_n)$ or censorship indicator vector $y_i = (y_1, y_2, ..., y_n)$ or vector are considered the observed data $\theta = (\eta, \beta)$.

First Step: Expectation

$$
\begin{aligned}
Q(\theta;\theta^k) &= E_{\theta^k}\left[l_c(x;\theta) \mid y,\delta,\theta^{k-1}\right] \\
&= n\ln\beta - n\beta\ln\eta + (\beta-1)\sum_{i=1}^{n} A_i^{(s)} - \frac{1}{\eta^\beta}\sum_{i=1}^{n} B_i^{(s)}
\end{aligned} \tag{5.47}
$$

where:

$$
A_i^{(s)} = E_{\theta^k}\left[\ln t_i \mid y,\delta\right] = \delta_i \ln y_i + (1-\delta_i)E_{\theta^k}\left[\ln t_i \mid t_i > y_i\right] \tag{5.48}
$$

$$
B_i^{(s)} = E_{\theta^k}\left[t_i^\beta \mid y,\delta\right] = \delta_i \eta_i^\beta + (1-\delta_i)E_{\theta^k}\left[t_i^\beta \mid y, t_i > y_i\right] \tag{5.49}
$$

The conditional probability is:

$$
f(y|y') = f(t_i|y_i) = \frac{f(t_i)}{1-F(y_i)}; t_i > y_i \quad \text{right censor} \tag{5.50}
$$

Making use of Weibull *PDF* function the expression became:

$$
\begin{aligned}
f(y \mid y') &= \frac{\frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{t_i}{\eta}\right)^\beta\right]}{1-\left[1-\exp-\left(\frac{y_i}{\eta}\right)^\beta\right]} \\
&= \frac{\frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{t_i}{\eta}\right)^\beta\right]}{\exp\left[-\left(\frac{y_i}{\eta}\right)^\beta\right]} \\
&= \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{t_i}{\eta}\right)^\beta\right]\exp\left[\left(\frac{y_i}{\eta}\right)^\beta\right] \\
&= \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta-1}\exp\left[\left(\frac{y_i}{\eta}\right)^\beta - \left(\frac{t_i}{\eta}\right)^\beta\right]
\end{aligned} \tag{5.51}
$$

To obtain the Expectation to A_i need to solve the following equation:

$$
\begin{aligned}
E_{\theta^{(k)}}\left[\ln t_i \mid t_i > y_i\right] &= \int_{y_i}^{\infty}\left\{\ln t_i \frac{\beta^{(k)}}{\eta^{(k)}}\left(\frac{t_i}{\eta^{(k)}}\right)^{\beta^{(k)}-1}\exp\left[\left(\frac{y_i}{\eta^{(k)}}\right)^{\beta^{(k)}} - \left(\frac{t_i}{\eta^{(k)}}\right)^{\beta^{(k)}}\right]\right\}dt_i \\
&= \exp\left(\frac{y_i}{\eta^{(k)}}\right)^{\beta^{(k)}}\int_{y_i}^{\infty}\left\{\ln t_i {\eta^{(k)}}^{\beta^{(k)}}\beta^{(k)} t_i^{\beta^{(k)}-1}\exp\left[-\left(\frac{t_i}{\eta^{(k)}}\right)^{\beta^{(k)}}\right]\right\}dt_i
\end{aligned}
$$

Make the substitution:

$$\left(\frac{t_i}{\eta}\right)^{\beta} = Z \Leftrightarrow t_i^{\beta} = \eta^{\beta} Z$$
$$t_i = \eta z^{\frac{1}{\beta}}$$
$$dt_i = \eta \frac{1}{\beta} z^{\frac{1}{\beta}-1}$$

can establish this result in a non-rigorous form, by using the fact that:

$$y \to \infty \Rightarrow Z \to \infty$$
$$Z_i = y_i \Rightarrow Z = \left(\frac{t_i}{\eta}\right)^{\beta}$$

And simplifying the following expressions (omitting the term K)

$$\begin{aligned}
E_{\theta^{(k)}}\left[\ln t_i \mid t_i > y_i\right] &= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} \ln(\eta Z^{\frac{1}{\beta}}) \eta^{-\beta} \beta \left(\eta Z^{\frac{1}{\beta}}\right)^{\beta-1} \exp(-Z) \eta \frac{1}{\beta} Z^{(\frac{1}{\beta}-1)} dZ \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} \ln(\eta Z^{\frac{1}{\beta}}) \eta^{-\beta} \eta^{\beta-1} Z^{(1-\frac{1}{\beta})} \exp(-Z) \eta Z^{(\frac{1}{\beta}-1)} dZ \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} \ln(\eta Z^{\frac{1}{\beta}}) \frac{1}{\eta} Z^{(1-\frac{1}{\beta})} \exp(-Z) \eta Z^{(\frac{1}{\beta}-1)} dZ \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} \ln(\eta Z^{\frac{1}{\beta}}) \exp(-Z) dZ \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} \left(\ln \eta + \frac{1}{\beta} \ln Z\right) \exp(-Z) dZ
\end{aligned}$$

And solve the integral:

$$
\begin{aligned}
E_{\theta^{(k)}}(A_i) &= \exp\left(\frac{y_i}{\eta}\right)^{\beta}\left\{\int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} e^{-Z}\ln\eta dZ + \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} e^{-Z}\frac{1}{\beta}\ln Z dZ\right\} \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta}\left\{-\ln\eta\int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} -\exp(-Z)dZ + \frac{1}{\beta}\int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty}\exp(-Z)\ln Z dZ\right\} \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta}\left\{-\ln\eta\left[e^{-Z}\right]_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} + \frac{1}{\beta}\left(\left[-e^{-Z}\ln Z\right]_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} - \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} -e^{-Z}\frac{1}{Z}dZ\right)\right\} \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta}\left\{\ln\eta\exp\left(-\frac{y_i}{\eta}\right)^{\beta} + \frac{1}{\beta}\left[\exp\left(-\frac{y_i}{\eta}\right)^{\beta}\ln\left(\frac{y_i}{\eta}\right)^{\beta} + \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} Z^{-1}\exp(-Z)dZ\right]\right\} \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta}\left\{\ln\eta\exp\left(-\frac{y_i}{\eta}\right)^{\beta} + \frac{1}{\beta}\left[\exp\left(-\frac{y_i}{\eta}\right)^{\beta}\ln\left(\frac{y_i}{\eta}\right)^{\beta} + \Gamma\left(0;\left(\frac{y_i}{\eta}\right)^{\beta}\right)\right]\right\} \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta}\ln\eta\exp\left(-\frac{y_i}{\eta}\right)^{\beta} + \frac{1}{\beta}\exp\left(\frac{y_i}{\eta}\right)^{\beta}\exp\left(-\frac{y_i}{\eta}\right)^{\beta}\ln\left(\frac{y_i}{\eta}\right)^{\beta} + \\
&\qquad \exp\left(\frac{y_i}{\eta}\right)^{\beta}\frac{1}{\beta}\Gamma\left(0;\left(\frac{y_i}{\eta}\right)^{\beta}\right) \\
&= \ln\eta + \frac{1}{\beta}\exp\left(\frac{y_i}{\eta}\right)^{\beta} + \exp\left(\frac{y_i}{\eta}\right)^{\beta}\frac{1}{\beta}\Gamma\left(0;\left(\frac{y_i}{\eta}\right)^{\beta}\right) \\
&= \ln\eta + \ln y_i - \ln\eta + \frac{1}{\beta}\exp\left(\frac{y_i}{\eta}\right)^{\beta}\Gamma\left(0;\left(\frac{y_i}{\eta}\right)^{\beta}\right) \\
&= \ln y_i + \frac{1}{\beta}\exp\left(\frac{y_i}{\eta}\right)^{\beta}\Gamma\left(0;\left(\frac{y_i}{\eta}\right)^{\beta}\right) \qquad (5.52)
\end{aligned}
$$

To obtain the Expectation to B_i need to solve the following equation:

$$
\begin{aligned}
E_{\theta^{(k)}}\left[t_i^{\beta}|t_i > y_i\right] &= \int_{y_i}^{\infty}\left\{t_i^{\beta^{(k)}}\frac{\beta^{(k)}}{\eta^{(k)}}\left(\frac{t_i}{\eta^{(k)}}\right)^{\beta^{(k)}-1}\exp\left[\left(\frac{y_i}{\eta^{(k)}}\right)^{\beta^{(k)}} - \left(\frac{t_i}{\eta^{(k)}}\right)^{\beta^{(k)}}\right]\right\}dt_i \\
&= \exp\left(\frac{y_i}{\eta^{(k)}}\right)^{\beta^{(k)}}\int_{y_i}^{\infty}\left\{t_i^{\beta^{(k)}}\eta^{(k)^{-\beta^{(k)}}}\beta^{(k)}t_i^{\beta^{(k)}-1}\exp\left[-\left(\frac{t_i}{\eta^{(k)}}\right)^{\beta^{(k)}}\right]\right\}dt_i
\end{aligned}
$$

Making the substitution

$$\left(\frac{t_i}{\eta}\right)^{\beta} = Z \Leftrightarrow t_i^{\beta} = \eta^{\beta} Z$$
$$t_i = \eta z^{\frac{1}{\beta}}$$
$$dt_i = \eta \frac{1}{\beta} z^{\frac{1}{\beta}-1}$$

And simplifying the following expressions (omitting the term K)

$$\begin{aligned}
E_{\theta^{(k)}}\left[t_i^{\beta}|t_i > y_i\right] &= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} \eta^{\beta} Z \eta^{-\beta} \beta \left(\eta Z^{\frac{1}{\beta}}\right)^{\beta-1} \exp(-Z) \eta \frac{1}{\beta} Z^{(\frac{1}{\beta}-1)} dZ \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} Z \beta \eta^{\beta-1} Z^{(1-\frac{1}{\beta})} \exp(-Z) \eta \frac{1}{\beta} Z^{(\frac{1}{\beta}-1)} dZ \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} Z \eta^{\beta-1} \exp(-Z) \eta dZ \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \eta^{\beta} \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} \exp(-Z) Z dZ
\end{aligned}$$

And solve the integral:

$$\begin{aligned}
E_{\theta^{(k)}}\left[t_i^{\beta}|t_i > y_i\right] &= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \eta^{\beta} \left\{ \left[-e^{-Z}.Z\right]_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} - \int_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} -e^{-Z} dZ \right\} \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \eta^{\beta} \left\{ 0 + e^{-\left(\frac{y_i}{\eta}\right)^{\beta}} \left(\frac{y_i}{\eta}\right)^{\beta} - \left[e^{-Z}\right]_{\left(\frac{y_i}{\eta}\right)^{\beta}}^{\infty} \right\} \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \eta^{\beta} \left\{ e^{-\left(\frac{y_i}{\eta}\right)^{\beta}} \left(\frac{y_i}{\eta}\right)^{\beta} + e^{-\left(\frac{y_i}{\eta}\right)^{\beta}} \right\} \\
&= \exp\left(\frac{y_i}{\eta}\right)^{\beta} \eta^{\beta} \left[e^{-\left(\frac{y_i}{\eta}\right)^{\beta}} \left(\left(\frac{y_i}{\eta}\right)^{\beta} + 1 \right) \right] \\
&= \eta^{\beta} \frac{(y_i)^{\beta}}{\eta^{\beta}} + \eta^{\beta} \\
&= y_i^{\beta} + \eta^{\beta}
\end{aligned} \tag{5.53}$$

Substituting 5.52 and 5.53 into 5.47 give the final result:

$$Q(\theta;\theta^k) = n\ln\beta - n\beta\ln\eta + (\beta-1)\sum_{i=1}^{n}\left[\delta_i \ln y_i + (1-\delta_i)\left(\ln y_i + \frac{1}{\beta}\exp\left(\frac{y_i}{\eta}\right)^{\beta}\Gamma\left(0;\left(\frac{y_i}{\eta}\right)^{\beta}\right)\right]\right.$$
$$-\frac{1}{\eta^{\beta}}\sum_{i=1}^{n}\left[\delta_i y_i^{\beta} + (1-\delta_i)\left(y_i^{\beta}+\eta^{\beta}\right)\right] \tag{5.54}$$

Derived in order to η:

$$\begin{aligned}\frac{\partial Q}{\partial \eta} &= 0 - \frac{n\beta}{\eta} + 0 - \left(-\beta\eta^{-\beta-1}\right)\sum_{i=1}^{n}\left[\delta_i y_i^{\beta} + (1-\delta_i)B_i\right]\\ &= -\frac{n\beta}{\eta} + \frac{\beta}{\eta^{\beta+1}}\sum_{i=1}^{n}\left[\delta_i y_i^{\beta} + (1-\delta_i)B_i\right]\end{aligned} \tag{5.55}$$

Derived in order to β:

$$\begin{aligned}\frac{\partial Q}{\partial \beta} &= \frac{n}{\beta} - n\ln\eta + \sum_{i=1}^{n}\left[\delta_i \ln y_i + (1-\delta_i)A_i\right] - \left(-\frac{\ln\eta}{\eta^{\beta}}\right)\sum_{i=1}^{n}\left[\delta_i y_i^{\beta} + (1+\delta_i)B_i\right]\\ &- \frac{1}{\eta^{\beta}}\sum_{i=1}^{n}\left[\delta_i y_i^{\beta}\ln y_i\right]\\ &= \frac{n}{\beta} - n\ln\eta + \sum_{i=1}^{n}\left[\delta_i \ln y_i + (1-\delta_i)A_i\right] + \sum_{i=1}^{n}\left\{\frac{\ln\eta}{\eta^{\beta}}\left[\delta_i y_i^{\beta} + (1+\delta_i)B_i\right] - \frac{1}{\eta^{\beta}}\delta_i y_i^{\beta}\ln y_i\right\}\end{aligned}$$

And simplifying

$$\begin{aligned}-\frac{n\beta}{\eta} + \frac{\beta}{\eta^{\beta+1}}A &= 0\\ -\eta^{\beta}n\beta + \beta A &= 0\\ \eta^{\beta}n\beta &= \beta A\\ \eta &= \left[\frac{\beta A}{\beta n}\right]^{\frac{1}{\beta}}\\ \eta &= \left(\frac{A}{n}\right)^{\frac{1}{\beta}}\end{aligned}$$

Calculate second derivative:

Derived in order to η

$$\frac{\partial^2 Q}{\partial \eta^2} = \frac{n\beta}{\eta^2} - \frac{\beta(\beta+1)}{\eta^{\beta+2}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} + (1-\delta_i)B_i\right]$$

Derived in order to β

$$\begin{aligned}
\frac{\partial^2 Q}{\partial \beta^2} &= -\frac{n}{\beta^2} + 0 + 0 + \sum_{i=1}^{n} \left\{ \ln\eta \left(-\frac{\ln\eta}{\eta^{\beta}}\right) \left[\delta_i y_i^{\beta} + (1-\delta_i)B_i\right] + \frac{\ln\eta}{\eta^{\beta}} \left[\delta_i y_i^{\beta} \ln y_i\right] \right\} \\
&- \sum_{i=1}^{n} \left\{ \left(-\frac{\ln\eta}{\eta^{\beta}}\right) \delta_i y_i^{\beta} \ln y_i + \frac{1}{\eta^{\beta}} \delta_i (y_i^{\beta} \ln y_i) \ln y_i \right\} \\
&= -\frac{n}{\beta^2} - \frac{(\ln\eta)^2}{\eta^{\beta}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} + (1-\delta_i)B_i\right] + \frac{\ln\eta}{\eta^{\beta}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} \ln y_i\right] + \frac{\ln\eta}{\eta^{\beta}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} \ln y_i\right] \\
&- \frac{1}{\eta^{\beta}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} (\ln y_i)^2\right] \\
&= -\frac{n}{\beta^2} - \frac{(\ln\eta)^2}{\eta^{\beta}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} + (1-\delta_i)B_i\right] + \frac{2\ln\eta}{\eta^{\beta}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} \ln y_i\right] - \frac{1}{\eta^{\beta}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} (\ln y_i)^2\right]
\end{aligned}$$

Derived in order to β and η

$$\begin{aligned}
\frac{\partial^2 Q}{\partial \eta \partial \beta} &= -\frac{n}{\eta} + \frac{1}{\eta^{\beta+1}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} + (1-\delta_i)B_i\right] + \left(-\frac{\ln\eta}{\eta^{\beta+1}}\right) \beta \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} + (1-\delta_i)B_i\right] \\
&+ \frac{\beta}{\eta^{\beta+1}} \sum_{i=1}^{n} \left(\delta_i y_i^{\beta} \ln y_i\right) \\
&= -\frac{n}{\eta} - \frac{1-\beta\ln\eta}{\eta^{\beta+1}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} + (1-\delta_i)B_i\right] + \frac{\beta}{\eta^{\beta+1}} \sum_{i=1}^{n} \left[\delta_i y_i^{\beta} \ln y_i\right]
\end{aligned}$$

Variance and covariance:

From the equation:

$$f(t_i|y_i) = \frac{\beta}{\eta} \left(\frac{t_i}{\eta}\right)^{\beta-1} \exp\left[\left(\frac{y_i}{\eta}\right)^{\beta} - \left(\frac{t_i}{\eta}\right)^{\beta}\right] \quad ; \quad t_i \geq y_i$$

and the logarithm

$$\begin{aligned}\ln f(t_i|y_i) &= \ln\left(\frac{\beta}{\eta}\right) + \ln\left(\frac{t_i}{\eta}\right)^{\beta-1} \ln\exp\left[\left(\frac{y_i}{\eta}\right)^{\beta} - \left(\frac{t_i}{\eta}\right)^{\beta}\right] \\ &= \ln\left(\frac{\beta}{\eta}\right) + \ln\left(\frac{t_i}{\eta}\right)^{\beta-1} + \left(\frac{y_i}{\eta}\right)^{\beta} - \left(\frac{t_i}{\eta}\right)^{\beta}\end{aligned}$$

Derived in order to η

$$\begin{aligned}\frac{\partial \ln f(t_i|y_i)}{\partial \eta} &= -\frac{1}{\eta} - \frac{1}{\eta^{\beta-1}}(\beta-1)\eta^{\beta-1-1} + \beta\left(\frac{y_i}{\eta}\right)^{\beta-1}\left(-\frac{y_i}{\eta^2}\right) - \beta\left(\frac{t_i}{\eta}\right)^{\beta-1}\left(-\frac{t_i}{\eta^2}\right) \\ &= -\frac{1}{\eta} - (\beta-1)\frac{\eta^{\beta-1}\eta^{-1}}{\eta^{\beta-1}} + \beta\left(\frac{y_i}{\eta}\right)^{\beta}\left(-\frac{1}{\eta}\right) - \beta\left(\frac{t_i}{\eta}\right)^{\beta}\left(-\frac{1}{\eta}\right) \\ &= -\frac{1}{\eta} - \frac{\beta-1}{\eta} + \frac{\beta}{\eta}\left(\frac{y_i}{\eta}\right)^{\beta} + \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta}\end{aligned}$$

Derived in order to β

$$\begin{aligned}\frac{\partial \ln f(t_i|y_i)}{\partial \beta} &= \frac{1}{\beta} + \frac{\left(\frac{t_i}{\eta}\right)^{\beta-1}\ln\left(\frac{t_i}{\eta}\right)}{\left(\frac{t_i}{\eta}\right)^{\beta-1}} + \left(\frac{y_i}{\eta}\right)^{\beta}\ln\left(\frac{y_i}{\eta}\right) - \left(\frac{t_i}{\eta}\right)^{\beta}\ln\left(\frac{t_i}{\eta}\right) \\ &= \frac{1}{\beta} + \ln\left(\frac{t_i}{\eta}\right) + \left(\frac{y_i}{\eta}\right)^{\beta}\ln\left(\frac{y_i}{\eta}\right) - \left(\frac{t_i}{\eta}\right)^{\beta}\ln\left(\frac{t_i}{\eta}\right)\end{aligned}$$

second derived in order to η

$$\begin{aligned}\frac{\partial^2 \ln f(t_i|y_i)}{\partial \eta^2} &= \frac{1}{\eta^2} + \frac{\beta-1}{\eta^2} + \frac{\beta}{\eta^2}\left(\frac{y_i}{\eta}\right)^{\beta} - \frac{\beta}{\eta}\beta\left(\frac{y_i}{\eta}\right)^{\beta-1}\left(-\frac{y_i}{\eta^2}\right) - \frac{\beta}{\eta^2}\left(\frac{t_i}{\eta}\right)^{\beta} + \frac{\beta}{\eta}\beta\left(\frac{t_i}{\eta}\right)^{\beta-1}\left(-\frac{t_i}{\eta^2}\right) \\ &= \frac{1}{\eta^2} + \frac{\beta-1}{\eta^2} + \frac{\beta}{\eta^2}\left(\frac{y_i}{\eta}\right)^{\beta} + \frac{\beta^2}{\eta}\left(\frac{y_i}{\eta}\right)^{\beta}\left(\frac{1}{\eta}\right) - \frac{\beta}{\eta^2}\left(\frac{t_i}{\eta}\right)^{\beta} - \frac{\beta^2}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta}\left(\frac{1}{\eta}\right) \\ &= \frac{1}{\eta^2} + \frac{\beta-1}{\eta^2} + \frac{\beta}{\eta^2}\left(\frac{y_i}{\eta}\right)^{\beta} + \frac{\beta^2}{\eta^2}\left(\frac{y_i}{\eta}\right)^{\beta} - \frac{\beta}{\eta^2}\left(\frac{t_i}{\eta}\right)^{\beta} - \frac{\beta^2}{\eta^2}\left(\frac{t_i}{\eta}\right)^{\beta} \\ &= \frac{1}{\eta^2} + \frac{\beta-1}{\eta^2} + \frac{\beta}{\eta^2}\left[\left(\frac{y_i}{\eta}\right)^{\beta} - \left(\frac{t_i}{\eta}\right)^{\beta}\right] + \frac{\beta^2}{\eta^2}\left[\left(\frac{y_i}{\eta}\right)^{\beta} - \left(\frac{t_i}{\eta}\right)^{\beta}\right]\end{aligned}$$

second derived in order to β

$$\begin{aligned}\frac{\partial^2 \ln f(t_i|y_i)}{\partial\beta^2} &= -\frac{1}{\beta^2} + 0 + \left(\frac{y_i}{\eta}\right)^{\beta} \ln\left(\frac{y_i}{\eta}\right) \ln\left(\frac{y_i}{\eta}\right) - \left(\frac{t_i}{\eta}\right)^{\beta} \ln\left(\frac{t_i}{\eta}\right) \ln\left(\frac{t_i}{\eta}\right) \\ &= -\frac{1}{\beta^2} + \left(\frac{y_i}{\eta}\right)^{\beta} \left(\ln\left(\frac{y_i}{\eta}\right)\right)^2 - \left(\frac{t_i}{\eta}\right)^{\beta} \left(\ln\left(\frac{t_i}{\eta}\right)\right)^2\end{aligned}$$

second derived in order to β and η

$$\begin{aligned}\frac{\partial^2 \ln f(t_i|y_i)}{\partial\beta\partial\eta} &= -\frac{1}{\eta} - \frac{1}{\eta}\left(\frac{y_i}{\eta}\right)^{\beta} - \frac{\beta}{\eta}\left(\frac{y_i}{\eta}\right)^{\beta} \ln\left(\frac{y_i}{\eta}\right) + \frac{1}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta} + \frac{\beta}{\eta}\left(\frac{t_i}{\eta}\right)^{\beta} \ln\left(\frac{t_i}{\eta}\right) \\ &= \frac{1}{\eta}\left[-1 - \left(\frac{y_i}{\eta}\right)^{\beta} + \left(\frac{t_i}{\eta}\right)^{\beta}\right] + \frac{\beta}{\eta}\left[-\left(\frac{y_i}{\eta}\right)^{\beta} \ln\left(\frac{y_i}{\eta}\right) + \left(\frac{t_i}{\eta}\right)^{\beta} \ln\left(\frac{t_i}{\eta}\right)\right]\end{aligned}$$

And the expected value

$$E\left[\frac{\partial^2 \ln f(t_i|y_i)}{\partial\eta^2} \mid t_i > y_i\right] = E\left[\frac{1}{\eta^2} + \frac{\beta-1}{\eta^2} + \frac{\beta}{\eta^2}\left(\frac{y_i}{\eta}\right)^{\beta} + \frac{\beta^2}{\eta^2}\left(\frac{y_i}{\eta}\right)^{\beta} - \frac{\beta}{\eta^2}\left(\frac{t_i}{\eta}\right)^{\beta} - \frac{\beta^2}{\eta^2}\left(\frac{t_i}{\eta}\right)^{\beta}\right]$$

5.8.2 Simulation EM to right censored

To simulate the method EM to right type I censored data with Weibull distribution the next algorithm have been developed:

Step 1: Define and model the firs step of method EM: Step E (expectation) as can see in 5.47

Step 2: Find and development the formulas to Step M (maximization)of method EM that result in equation 5.54

Step 3: Generate U_i random uniform (0,1)

Step 4: Generate t_i and δ_i from *step 3* and repeat for n times (the dimension of the sample)

Step 5: Define the initial values θ_k that will begin the algorithm iteration

Step 6: Using the equations from *step 2* find θ_{k+1}

Step 7: Compare $\epsilon = L(\theta_{k+1}) - L(\theta_{k+1})$ and stop if the value $\epsilon > 0$ are acceptable, otherwise repeat Step 6.

Two different algorithms have been made: the first one was made in *python* and the second was made in *R*.

The algorithm in *python* have two principals routines: the first is the definition of the two functions (algorithm 7), and the second is the cycle that use the numerical solution of the equation 5.54 by the function *nsolve* from the package system - *Sympy*.

The algorithm in *R* have also two principals routine: the functions that represents symbolically the formulas from 5.54 (algorithm 8) and the routine optimize and solve the equation 5.54 using the function *optim* from R software.

```
1
2  def funs (y):
3      b,ni=y
4      for i in range(0,n):
5          xx=Vr[i]*Yi[i]**b+(1-Vr[i])*(Yi[i]**(bk)+(nk)**(bk))
6          lti1.append(xx)
7      f1=-n*b/ni+b/(ni**(b+1))*sum(lti1)
8      for i in range(0,n):
9          tt=mp.gammainc(0,(Yi[i]/nk)**bk)
10         rr=float(tt)
11         xx=Vr[i]*sp.log(Yi[i])+(1-Vr[i])*(sp.log(Yi[i])
12 +1/(bk)*sp.exp((Yi[i]/nk)**bk)*rr)
13         lti2.append(xx)
14     for i in range(0,n):
15             xx=sp.log(ni)/ni**b*(Vr[i]*Yi[i]**b
16 +(1-Vr[i])*(Yi[i]**(bk)+(nk)**(bk)))-1/ni**b*Vr[i]*Yi[i]**b
17 *sp.log(Yi[i])
18             lti3.append(xx)
19     f2=n/b-n*sp.log(ni)+sum(lti2)+sum(lti3)
```

Algorithm 7: EM to right censored with python - define function (partial)

```
# function that define calculate the two parameters
ols.lf<-function(theta,dx,Vr){
  b<-theta[1]
  ni<-theta[2]
  ss <- Vr*log(dx)-(1-Vr)*(log(dx)+1/(bk)
        *exp(dx/nk)^bk*gamma_inc(0,((dx/nk)^bk)))
  tt <-  Vr*dx^b+(1-Vr)*(dx^bk+nk^bk)
  logl<- n*log(b)-n*b*log(ni)+(b-1)*sum(ss)-1/(ni^b)*sum(tt)
```

Algorithm 8: EM to right censored with R - define function (partial)

```
for (i in 1:m){
# Formula to calculate the n(th) shape and scale estimated
  dx <- rweibull(n,sh,sc)
  Vr <- rbinom(n, 1, cen)
  cr=0; zz=bk
  condition = 10

  while (condition == 10) {
      p<-optim(c(bk,nk),ols.lf,method="BFGS",dx=dx,Vr=Vr)

      #retrive values to loop
        tbk=abs(bk- p$par[1])
        bk <-p$par[1]; nk <- p$par[2]
        cr=cr+1; zz[cr]=bk
          if (tbk<0.01 ) {condition =0} else { }
          if (cr==15 ) {condition =0} else { }
  }
  f1data <- cbind(f1data,zz)
```

Algorithm 9: EM to right censored with R - cycle (partial)

A huge number of simulations was made, but to summarize how the algorithm work, only the results of *R* software is present. The algorithms made in *python* have found very difficult to tuning the initial values and, in a great number of times, the value of convergence was to far from the theoretical value.

The results from R software are more precise, and the time and consuming of the computer processor is much lesser than with python software.

β= 1	η= 10	N. Iterations
1.1992	11.0837	8
1.0283	11.9851	4
1.1214	8.6873	2
1.0966	11.2708	2
1.0959	13.6491	1
1.0195	11.2788	3
1.4692	14.0542	12
1.1658	13.1808	8
0.9013	10.8018	4
1.0311	10.8472	4

Table 5.4: Simulation EM Method, Right type I, Weibull, $\beta = 1$, $\eta = 10$, $n = 100$

The basic procedure of the EM algorithm was applied to determine the solutions of the equation that derives from the maximum likelihood method, in particular in the case of right censored data. The EM algorithm has several properties and advantages that stand out from other iterative algorithms, such as: - the required analytical work is simpler than with other methods, since it is only necessary to maximize the conditional expected value of the log-likelihood and is relatively easy to program and implement.

The table 5.4 present the result from one of the multiple simulations,that have been done, with different shapes factor and censoring time.

In the analysis of the table for $beta = 1$ e $\eta = 10$ the number of iterations is low and the evolution can be better see in the figure 5.1. The estimation is much more precise to the shape factor β than to the scale factor η.

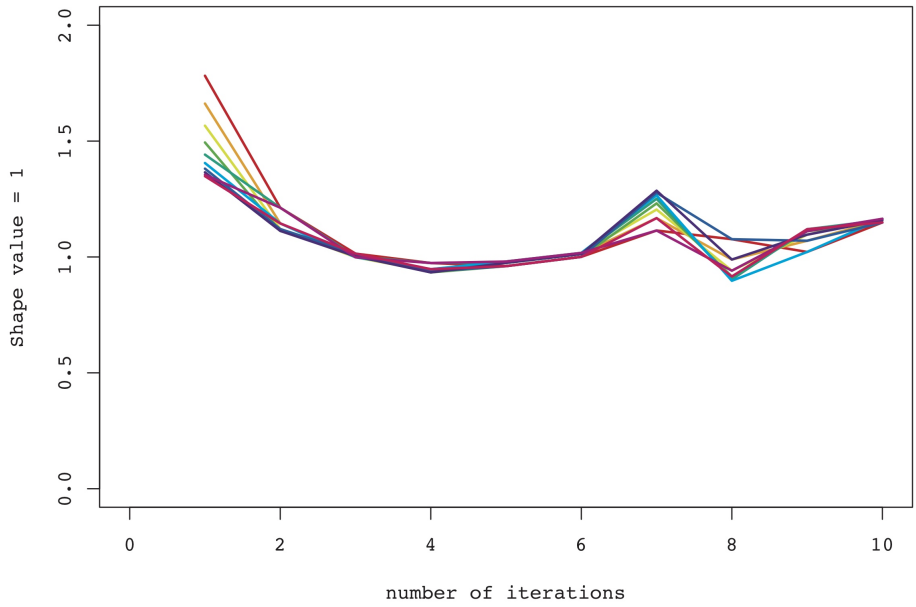


Figure 5.1: Iterations and evolution *EM* method

The figure 5.1 show the results and the evolution of convergence of the algorithm from ten cycles with minimum of ten iterations.

The simulation show the algorithms developed converged quickly but in other situations the EM algorithm may converge slowly, even in some seemingly simple problems, and in problems in which there is a lot of incomplete information. However the EM algorithm, like other methods, does not guarantee the convergence to the global maximum when there are several local maximum. The estimate obtained depends on the initial solution.

In this section, it was possible to verify the development of the EM algorithm with a numerical method to determine the solutions of the equations that derives from the maximum likelihood method, in particular in the case of the presence of right censored data.

Chapter 6

Reliability and simulation of systems

The subject of reliability and simulation of systems and the structural relationship between a system and its components is very important in the field of reliability. A comprehensive discussion of reliability theory can be found in Barlow and Proschan (1975) and in Kaufmann et al. (1977). Reliability Block Diagrams - RBD can be defined as network of blocks describing the function of the system with logical connections of components needed to produce a specified system function.

The theory and the reliability functions of the linear consecutive systems are derived and algorithms and simulation are carried out with a new and original approach.

The component importance measure can be a very important tool to help to optimize maintenance resources, inspection plans or improve better preventive maintenance tasks.

6.1 Systems and Reliability Block Diagrams

To model the system (equipment, line production, etc...) there is a lots of tools, two of them is RBD and FT - fault tree. Sometimes, the two tools give the same result and can convert the fault tree to reliability block diagram, and vice-versa, this could happen when fault tree is only OR-gates, or AND-gates and in RBD is only series-parallel, not complex system. Barlow and Proschan (1975) present an exhaustive description of the theory of RBD and more recently Rausand and Høyland (2004). If the system has more than one function, each function must be considered individually, and different diagram need to be made for each system function. The system is fixed in one moment of time; the present state of the system is assumed to depend only on the present states of the components.

The connection through a block, in RBD means that the component is functioning. The series structure is equivalent to a fault tree where all the events are connected through an OR-gate. The TOP event occurs if either component 1 or component 2 or component 3 or .... component n fails. The parallel structure can be represented as a fault tree where all the basic event are connected through an AND-gate. The TOP event occurs if component 1 and component 2 and component 3 ... and component n fail.

In the construction of the fault tree there is a search for all potential causes of a specified system failure. The FT - fault tree will give to the analyst a better understanding of the causes of failure and this potential and risks. If the analysis is done in the design or development phase, the researcher may redesign the operation of the system and take actions to improve and eliminate potential hazard.

A Reliability Block Diagram is developed in terms of functions. Normally not have in account safety and auxiliary functions and components used to protect equipment, people or the environment. Reliability Block Diagrams can be used for repairable systems and for non repairable systems or components.

A Fault Tree will be converted in Reliability Block Diagram for qualitative and quantitative analyses. For further evaluations is often more natural to base analysis on a reliability block diagram and this is the main reason to chosen the focus on reliability block diagram in the rest of chapter.

6.1.1 System of components

A system that is composed with n components will be classified a system of order n. The component are to be numbered consecutively from 1 to n.

Let $C = \{c_1, c_2, ..., c_n\}$ be the set of all components, where c_i is the ith component, and n is the number of components in the system.

Let x_i be the state of component c_i the system can be in one and only one of two states, that is either functioning or failed.

In chapter 6 there is only two states to systems and to components: a functioning state and a failed state.

To indicate the state of the ith component a binary indicator variable x_i to component i is assigned:

$$x_i = \begin{cases} 1 & \text{if component } i \text{ is functioning,} \\ 0 & \text{if component } i \text{ is failed.} \end{cases}$$

for i=1,...,n, where n is the number of the components in the system.

The number of components n in the system is called the *order* of the system.

The joint performance of all components in the system can be indicated by vector $X = (x_1, x_2, ..., x_n)$ called a state vector.

Similarly, the binary variable ϕ indicates the state of the system:

$$\phi = \begin{cases} 1 & \text{if system } i \text{ is functioning,} \\ 0 & \text{if system } i \text{ is failed.} \end{cases}$$

The term *binary variable* will refer to a variable taking on the values 0 or 1.

The state of system is determined completely by the states of the components, so that may write

$$\phi = \phi(x), \quad where \quad x = (x_1, ..., x_n).$$

The function $\phi(x)$ is called the *structure function* of the system. A knowledge of the structure function is equivalent to a knowledge of the structure of the system.

6.1.2 Series structure

A system that is functioning if and only if all of its n components are functioning is called a *series structures*. The corresponding reliability block diagram is shown in figure 6.1.

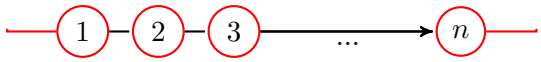


Figure 6.1: Series structure diagram

A *series structure* functions if and only if each component functions. The structure function is given by

$$\phi(X) = \prod_{i=1}^{n} x_i = min(x_i, ..., x_n).$$

6.1.3 Parallel structure

A system that is functioning if at least one of its n components is functioning is called a *parallel structure*. The corresponding reliability block diagrams is shown in figure 6.2.

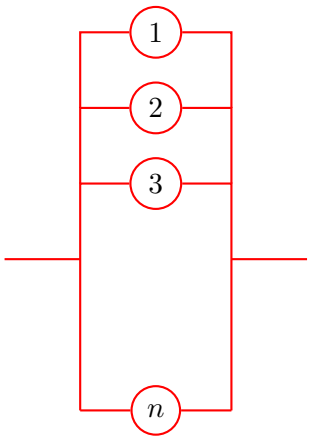


Figure 6.2: Parallel structure diagram

A *parallel* structure functions if and only if at least one component works. The structure function is given by

$$\phi(X) = \coprod_{i=1}^{n} x_i = max(x_i, ..., x_n).^{1}$$

6.1.4 The definition *k-out-of-n* structure

A system that is functioning if and only if at least k of the n components functioning is called a *k-out-of-n* structure functions

The structure function *k-out-of-n* structure can be written

$$\phi(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i \geq k, \\ 0 & \text{if } \sum_{i=1}^{n} x_i \leq k. \end{cases}$$

or equivalently,

$$\phi(x) = \prod_{i=1}^{n} x_i \quad for \quad k = n,$$

while

$$\phi = (x_1...x_k) \amalg (x_1...x_{k-1}x_{k+1}) \amalg ... \amalg (x_{n-k+1}...x_n)$$

$$\equiv max\left\{(x_1...x_k) \amalg (x_1...x_{k-1}x_{k+1}) \amalg ... \amalg (x_{n-k+1}...x_n)\right\}$$

for $1 \leq k \leq n$, where every choice of k out of the n x's appears exactly.

A parallel structure is a *1-out-of-n* structure and a series structure is an *n-out-of-n* structure.

The system have component irrelevant when, if the state of that component change but the change the state of the system doesn't, for all possible states of all components in the system.

When a system structure has irrelevant components, it may be possible to simplify it by omitting those components irrelevant to its performance. This structure is called reducible

The structure is called irreducible when all components are relevant. Coherent structure is an irreducible structure with monotone structure function.

A monotone structure is a structure function ϕ, is called monotone if $\phi(y) \geq \phi(x)$ for all $Y \supseteq X$, where $Y \supseteq X \Leftrightarrow y_i \geq x_i$ for all i

---

[1] Notation:

$$\coprod_{i=1}^{n} x_i = 1 - \prod_{i=1}^{n}(1 - x_i), \quad and \quad x_1 \amalg x_2 = 1 - (1 - x_1)(1 - x_2)$$

6.1.5 Coherent structures

The system with a monotone structure function is called semi-coherent. A semi-coherent system having relevant components is then called a coherent system,

There are only two semi-coherent structures that are not coherent: is $\phi(0) = 0$, which fails for every state of its components, and the structure $\phi(1) = 1$, which performs for every state of its components.

Examples of typical coherent systems are:

1. Series system of order n with structure function $\phi(x) = \prod_{i=1}^{n} x_i$,
2. A parallel system of order n with structure function $\phi(x) = \coprod_{i=1}^{n}(1-)x_i)$,
3. A *k-out-of-n* system with structure function $\phi(X) = 1 \quad if \quad s(X) \geqslant k$ and 0 otherwise, and
4. A parallel-series (series-parallel) system: the system consisting of a parallel (series) arrangement of series (parallel) subsystem.

A physical system would be poorly designed if improving the performance of a component (that is, replacing a failed component by a functioning component) caused the system to deteriorate.

Assume that, the system will not run worse than before if we replace a component in a failed state with one that is functioning. This is obviously the same as requiring that the structure function shall be non-decreasing in each of its arguments.

A system of components is said to be coherent if all its components are relevant and the structure function is non-decreasing.

6.1.6 Structures represented by paths and cuts

A structure of order n consists of n components numbered from *1* to n. The set of components is denoted by

$$c = \{1, 2, ...n\}$$

A path set P is a set of component in C which by functioning ensures that the system is functioning. A path set is said to be minimal if it cannot be reduced without loosing its status as a path set.

A critical path vector for components i is a state vector $(1_i, x)$ such that

$$\phi(1_i, x) = 1 \quad while \quad \phi(0_i, x) = 0$$

This is equivalent to requiring that

$$\phi(1_i, x) - \phi(0_i, x) = 1$$

In other words, given the states of the other components $(\cdot_i, x)$, the system is functioning if and only if component i is functioning. It is therefore natural to call $(1_i, x)$ a *critical* path vector for component i.

A critical path set $C(1_i, x)$ corresponding to the critical path vector $(1_i, x)$ for component i is defined by:

$$C(1_i, x) = \{i\} \cup \{j; x_j = 1, j \neq i\}$$

The total numbers of critical path sets (path vectors) for component i is

$$\eta_\phi(i) = \sum_{(\cdot_i, x)} [\phi(1_i, x) - \phi(0_i, x)]$$

Since the $x_j's$ are binary variables and thus can take only two possible values, *0* and *1*, the total number of state vectors $(\cdot, x) = (x_1, \cdots, x_{i-1}, \cdot, x_{i+1}, \cdots, x_n)$ is 2^{n-1}.

A cut set K is a set of components in C which by failing causes the system to fail. A cut set is said to be minimal if it cannot be reduced without loosing its status as a cut set.

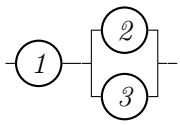


Figure 6.3: Reliability block diagram

Example 6.1.1:
Consider the reliability block diagram in fig. 6.3. The component set is $c = \{1, 2, 3\}$.
In this case the minimal path sets are

$$P_1 = \{1, 2\} \quad and \quad P_2 = \{1, 3\}$$

while the minimal cut sets are

$$K_1 = \{1\} \quad and \quad K_2 = \{2, 3\}$$

The association of a binary function with the arguments x_i and $i \in P_j$ with the jth minimal path set P_j of a coherent structure ϕ is represent by:

$$\rho_j(x) = \prod_{i \in P_j} x_i, \tag{6.1}$$

if all components in the jth minimal path set function takes the value 1 , and 0 otherwise.

The structure ρ_j as the jth *minimal path series structure*, (J=1,...,p) where p is the number of minimal path sets of ϕ and ρ_j is the structure function of a series arrangement of the components of the jth path set.

The structure are working if and only if at least one of the minimal path structures is work, follow identity is true representing the structure as a parallel arrangement of the minimal path series structures:

$$\phi(x) \equiv \coprod_{j=1}^{P} \rho_j(x) \equiv \prod_{j=1}^{P} [1 - \rho_j(x)], \tag{6.2}$$

The structure may be interpreted as a parallel structure of the minimal path series structures. Combining equation 6.2 with equation 6.1 we get:

$$\phi(x) = \coprod_{j=1}^{P} \prod_{i \in P_j} x_i \tag{6.3}$$

The association of a a binary function with arguments $x_i, i \in K_j$ with the jth minimal cut set K_j of a coherent structure ϕ, is represent by:

$$\kappa_j(x) = \coprod_{i \in K_j} x_i, \tag{6.4}$$

if all the components in the jth minimal cut set fail,which takes the value 0 and 1 otherwise. The structure κ_j as the jth *minimal parallel cut structure*, (J=1,...,K) where K is the number of minimal cut sets of ϕ and κ_j is the structure function of a parallel arrangement of the components of the jth cut set.

The structure fails if and only if at least one of the minimal cut structures fails, the follow identity is true:

$$\phi(x) = \prod_{j=1}^{K} \kappa_j(x) \tag{6.5}$$

representing the structure as a series arrangement of the minimal cut parallel structures.

The structure may be interpreted as a series structure of the minimal cut parallel structure. By combining eq. 6.4 and 6.5 we get

$$\phi(x) = \prod_{j=1}^{K} \coprod_{i \in K_j} x_i \tag{6.6}$$

Example 6.1.2 (The Bridge Structure)**:**
Consider a bridge structure such as that given by the physical network in figure 6.4
The minimal path sets are

$$P_1 = \{1, 4\}, \quad P_2 = \{2, 5\}, \quad P_3 = \{1, 3, 5\}, \quad P_4 = \{2, 3, 4\}$$

The minimal cut sets are:

$$K_1 = \{1, 2\}, \quad K_2 = \{4, 5\}, \quad K_3 = \{1, 3, 5\}, \quad K_4 = \{2, 3, 4\}$$

Using the equation 6.2 and the minimal path sets, the bridge structure may be represent as a parallel-series diagram.

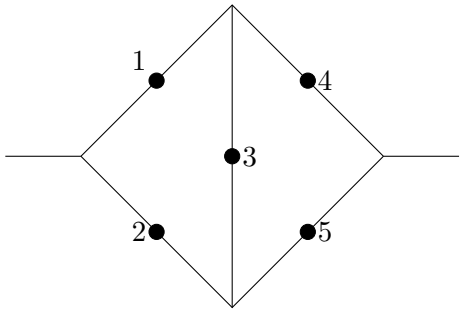


Figure 6.4: Bridge Structure Diagram

$$
\begin{aligned}
\rho_1 &= x_1 \cdot x_3 \cdot x_5 \\
\rho_2 &= x_2 \cdot x_3 \cdot x_4 \\
\rho_3 &= x_1 \cdot x_4 \\
\rho_4 &= x_2 \cdot x_5
\end{aligned}
\tag{6.7}
$$

accordingly, the structure function may be written:

$$
\begin{aligned}
\phi(x) = \coprod_{j=1}^{4} \rho_j(x) &= 1 - \prod_{j=1}^{4}(1 - \rho(x)) \\
&= 1 - (1 - \rho_1(x))(1 - \rho_2(x))(1 - \rho_3(x))(1 - \rho_4(x)) \\
&= 1 - (1 - x_1x_4)(1 - x_2x_5)(1 - x_1x_3x_5)(1 - x_2x_3x_4) \\
&= x_1x_4 + x_2x_5 + x_1x_3x_5 + x_2x_3x_4 - x_1x_3x_4x_5 - x_1x_2x_3x_5 \\
&- x_1x_2x_3x_4 - x_2x_3x_4x_5 - x_1x_{2x4}x_5 + 2x_1x_2x_3x_4x_5
\end{aligned}
\tag{6.8}
$$

Similarly, using equation 6.5 and the minimal cut sets, we may represent the bridge as a series-parallel structure:

$$
\begin{aligned}
\kappa_1 &= 1 - (1 - x_1)(1 - x_2) \\
\kappa_2 &= 1 - (1 - x_4)(1 - x_5) \\
\kappa_3 &= 1 - (1 - x_1)(1 - x_3)(1 - x_5) \\
\kappa_4 &= 1 - (1 - x_2)(1 - x_3)(1 - x_4)
\end{aligned}
$$

Given the minimal cut (or minimal path) sets, any coherent system can be represented as in the above "circuit" diagrams.

6.1.7 Pivotal decomposition

The following *pivotal decomposition* of the structure function, fundamental tool in carrying through inductive proofs, is made:

The following identity holds for any structure function ϕ of order n

$$\phi(x) = x_i\phi(1_i, x) + (1 - x_i)\phi(0_i, x) \quad \text{for all} \quad x(i = 1, ..., n) \tag{6.9}$$

The equation 6.9 permit us to express a structure function of order n in terms of structure functions of order *n-1*. By repeat the formulations, we obtain the representation:

$$\phi(x) = \sum_y \prod_{j=1}^{n} x_j^{y_j}(1 - x_j)^{1-y_j}\phi(y), \tag{6.10}$$

where the sum is extended over all binary vectors y of order n, and $0^0 \equiv 1$.

A structure function $\phi(X)$ of a system of order n has a property that it can be expressed in terms of the structure functions of systems of order *n-1*, which is called a pivotal decomposition and is written as

$$\phi(x) = x_i\phi(1_i, x) + (1 - x_i)\phi(0_i, x)$$

for all $(\cdot_i, x)$ and $i = 1, ..., n$

The structure function ϕ of a coherent system can therefore be written in the form

$$\phi(x) = x_i\left[\phi(1_i, x) - \phi(0_i, x)\right] + \phi(0_i, x) = x_i\delta_i(x) + \mu_i(x)$$

where $\delta_i(x) = \phi(1_i, x) - \phi(0_i, x) = \partial\phi(x)/\partial x_i$ and $\delta_i(x)$ as well as $\mu_i(x)$ do not depend on the state x_i of component c_i.

Repeated application of this method will permit us to explicitly express the structure function $\phi(X)$ in terms of the state of its components. This application can be written as follow:

$$
\begin{aligned}
\phi(X) &= x_i\phi(1_i, x) + (1 - x_i)\phi(0_i, x) \\
&= x_1\phi(1, x_2, x_3, ..., x_n) + (1 - x_1)\phi(0, x_2, x_3, ..., x_n) \\
&= x_1\left[x_2\phi(1, 1, x_3, ..., x_n) + (1 - x_2)\phi(1, 0, x_3, ..., x_n)\right] \\
&+ (1 - x_1)\left[x_2\phi(0, , 1, x_3, ..., x_n) + (1 - x_2)\phi(0, 0, x_3, ..., x_n)\right] \\
&\cdots \\
&= \sum_y \prod_{j=1}^{n} x_j^{y_j}(1 - x_j)^{1-y_j}\phi(y),
\end{aligned} \tag{6.11}
$$

where the summation is taken over some possible values of vector y such that $\phi(y) = 1$. Equation 6.11 is called *Barlow-Proschan* representation of the structures.

The representation of a structure function using pivotal decomposition is a basic way of expressing the structure in terms of the states of system component, but for coherent systems, this formula gives too many terms that, in fact, can be cancelled out in the process of simplification.

6.1.8 Reliability function in coherent structures

Let the component state X_i be a Bernoulli random variable ($P(X_i = 1) = p_i$ and $P(X_i = 0) = q_i$, where $q_i = 1 - p_i$. Then $P(X_i = 1) = p_i$ is called the reliability of component c_i and $i = 1, 2, ..., n$. The corresponding system reliability is given by:

$$R_\phi(P) = P\{\phi(X) = 1|p\} = E[\phi(X)|p], \quad p = p_1, p_2, ..., p_n \tag{6.12}$$

R_ϕ is called the reliability based on the structure function ϕ. By using the assumption of component independence, the reliability R_ϕ is whole determined by component reliabilities $(p_1, p_2, ..., p_n)$.

R_ϕ can be written as $R_\phi(P)$ and in case of $(p_1 = p_2 = ... = p_n = p)$, the reliability function can be written as $R_\phi(p)$ from a common component reliability p.

When component c_i is functioning, the reliability function of a system is:

$$R_\phi(1_i, p) = P\{\phi(1_i, X) = 1|p\}; \tag{6.13}$$

and the reliability function of a system given component c_i fails is:

$$R_\phi(0_i, p) = P\{\phi(0_i, X) = 1|p\}; \tag{6.14}$$

where $p = (p_1, ..., p_{i-1}, p_{i+1}, ..., p_n)$.

The reliability $R_\phi(P)$ can also be expressed in the pivotal decomposition, This is written as:

$$R_\phi(P) = p_i R_\phi(1_i, p) + (1 - p_i) R_\phi(0_i, p). \tag{6.15}$$

Exact system reliability can be computed using the structure function $\phi(X)$.

$$= E(\coprod_{i=1}^{t} \coprod_{\{j:z_{ij}=0\}} X_j) = E(\prod_{i=1}^{r} \coprod_{\{j:w_{ij}=0\}} X_j), \tag{6.16}$$

where z_{ij} is the jth element of minimal-path vector z_i; and where w_{ij} is the jth element of minimal-cut vector w_i. And assume independent components.

For structure function in *"reduced"* form, that is, a structure function having no power of x_i greater than *1*, the reliability function is simply the structure function ϕ evaluated on p.

$$R_\phi(X) = \sum_x \prod_{i=1}^{n} p_i^{x_i} (q_i)^{1-x_i} \phi(X), \tag{6.17}$$

where x_i is the value of the ith elements in path vectors X.

If a system has a common component reliability p, $\Re_\phi(P)$ can be computed by

$$R_\phi(X) = \sum_{i=1}^{n} P_i p^i q^{n-i}, \tag{6.18}$$

where P_i denotes the number of path vectors X of size i. For example, the reliability function of *k-out-of-n* systems having a common component reliability p is

$$R_\phi(p, n, k) = \sum_{i=k}^{n} \binom{n}{i} p^i q^{n-i}, \tag{6.19}$$

6.1.9 Reliability function with life distribution

Finally, let us consider the expression of reliability function when a system, as well its components, has life distribution.

Let $F_i(t)$ be a life distribution of component c_i and let

$$X_i(t) = \begin{cases} 1 & \text{if } c_i \text{ is functioning until time t} \\ 0 & \text{if } c_i \text{ is failed before time t.} \end{cases}$$

Then, the reliability of component c_i at time t is

$$P[X_i(t) = 1] = E[X_i(t)] = S_i(t) = 1 - F_i(t), \tag{6.20}$$

and the system reliability at time t is

$$R_\phi[S(t) = 1] = P\{\phi[X(t)] = 1\} = E_\phi[X_i(t)]. \tag{6.21}$$

6.2 Simulation for coherent system: k-out-of-n

To made a simulation Monte Carlo of complex systems *k-out-of-n*, using Weibull distribution the follow algorithm have been developed:

Step 1: Define the function of the structure of *k-out-of-n* system

Step 2: Calculate the time to censoring - t_c with the parameters of distribution chosen

Step 3: Generate t_i from random distribution function

Step 4: Compare the time t_i with T_c to each component and give $x_i = 0$ if are above or $x_i = 1$ below the t_c

Step 5: With x_i and structure function calculate if the system are working or not

Step 6: Repeat for$M$ times (the dimension of the cycle simulation)

Step 7: Calculate the reliability: the number of times that the system are working for the number of samples M

The program has been written in *R* software, and in the beginning is define the function structure and the rest of parameters (number of simulations, etc...). After that, is applied the loop *"for"* to make the cycle where reliability is calculated- it's the Monte Carlo simulation core. The difference between *2-out-of-3* and the bridge example *2-out-of-5* is the structure function of the system. This is define in the program by a function with the name *str fun*.

```
str_fun<-function(x1,x2,x3,x4,x5){
  res=x1*x4+x2*x5+x1*x3*x5+x2*x3*x4-x1*x3*x4*x5-x1*x2*x3*x5
-x1*x2*x3*x4-x2*x3*x4*x5-x1*x2*x4*x5+2*x1*x2*x3*x4*x5

  return(res)
}

simul_fun<-function(m){
  cr=0
```

Algorithm 10: Simulation for k-out-of-n function *str fun*

6.2.1 Simulation of coherent system: 2-out-of-3

To all components it's used the same Weibull parameters: the shape parameter β have the values *0,5;1;1,5 and 2*; the scale parameter is $\eta = 10$, for all components and simulations. The simulation is made for different number of samples to verified the impact of the number of samples for each simulation. Another interest characteristic is to simulate reliability with different censored data, in this case it's used 5%,10%, 20% and 30%.

	$\mathbf{C}_{5\%}$				$\mathbf{C}_{10\%}$				$\mathbf{C}_{20\%}$				$\mathbf{C}_{30\%}$			
Sample	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2
10	1	0.7	0.8	0.6	1	1	0.7	0.8	1	1	1	0.7	1	1	0.9	0.7
100	0.94	0.94	0.82	0.69	0.99	0.99	0.93	0.76	1	1	0.96	0.81	1	1	0.99	0.88
500	0.91	0.88	0.82	0.73	1	0.98	0.93	0.76	1	0.99	0.95	0.86	1	1	0.98	0.9
1000	0.93	0.88	0.81	0.74	0.99	0.97	0.89	0.78	1	1	0.96	0.82	1	1	0.98	0.86
2000	0.94	0.89	0.8	0.75	0.99	0.98	0.9	0.78	1	1	0.95	0.82	1	1	0.99	0.86

Table 6.1: Simulation 2-out-of-3, Weibull $(\beta, \%C, n)$, $\eta = 10$

The results are explicit in the table 6.1. From the analysis of the table, it's possible to see that with the increase of the sample number the value of reliability stabilized at a certain value. The increase of the shape factor β, the value of reliability decreases, however the most unexpected evolution data , is the growth of reliability with the increase of censored data, when it was expected to decrease. One possible explanation is that with the increase of censorship, the system itself ends up better, vanish the faults and therefore give a globally less damage.

Test with other parameters and also make a comparative test with different parameters for each component will be recommend.

6.2.2 Simulation for 2-out-of-5 Bridge structure

To all components and similar with *2-out-of-3* it's used the same Weibull parameters: the shape parameter β have the values *0,5;1;1,5 and 2* and the the scale parameter is $\eta = 10$. The simulation is made for different number of samples to verified the impact of the number of samples for each simulation.

The results are explicit in the table 6.2 and compare with the simulation of reliability *2-out-of-3* system, the evolution and the pattern are the same. The absolute values of reliability for each simulation is very close. Although, this structure is more complex and have almost the same redundancy as the simple *2-out-of-3* system.

	$\mathbf{C}_{5\%}$				$\mathbf{C}_{10\%}$				$\mathbf{C}_{20\%}$				$\mathbf{C}_{30\%}$			
Sample	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2
10	0.9	0.9	0.7	0.7	1	1	0.8	0.6	1	1	1	0.8	1	1	1	0.9
100	0.92	0.89	0.79	0.71	1	0.95	0.88	0.81	1	0.99	0.97	0.88	1	1	1	0.83
500	0.94	0.9	0.82	0.74	0.99	0.98	0.91	0.82	1	1	0.96	0.85	1	1	0.98	0.89
1000	0.93	0.9	0.83	0.77	0.99	0.97	0.91	0.8	1	1	0.95	0.86	1	1	0.99	0.91
2000	0.93	0.9	0.83	0.76	1	0.98	0.91	0.8	1	1	0.97	0.84	1	1	0.99	0.88

Table 6.2: Simulation 2-out-of-5, Weibull $(\beta, \%C, n)$, $\eta = 10$

6.3 Linear Consecutive k-out-of-n systems

Consecutive *K-out-of-n* systems can be categorized into consecutive *k-out-of-n:G* and consecutive *k-out-of-n:F* systems.

A Consecutive ***k-out-of-n:G*** system is an n component system that functions whenever at least k consecutive components are functioning.

A consecutive ***k-out-of-n:F*** is an n component system that fails whenever at least k consecutive component are failed.

Such system can either be a linear system, where all components are arranged linearly, or be a circular system, where all components are arranged circularly.

The consecutive K-out-of-n:F system was introduced by Kontoleon (1980), Chiang and Niu (1981) and explain the relevance of such a system to telecommunication and oil pipeline systems. The application to street light systems and microwave tower systems are discussed by Bollinger and Salvia (1982).

The structure and reliability functions for both systems follow the notion and development made by Zakaria (1989).

6.3.1 Structure of linear consecutive k-out-of-n:G systems

The minimal-path representation is used to find the structure function of the linear consecutive k-out-of-n:G system, where $2 \leqq k \leqq n$. The number of minimal-path vectors for this system is equal to $n-k+1$ possibilities for placing the K consecutive *1's* on n possible locations. So, the set of minimal-path vectors for linear consecutive *k-out-of-n:G* systems is:

$$Z = \{Z_i\}_{i=1}^{n-k+1} = \{(1_k, 0_{n-k}), (0, 1_k, 0_{n-k-1}), ..., (0_{n-k}, 1_k)\}. \tag{6.22}$$

where 0_j is the j-dimensional zero vector *(0,0,...,0)* and 1_j is the j-dimensional unit vector *(1,1,...,1)*.

Based on the minimal-path vector representation the structure function for the linear consecutive *k-out-of-n:G* systems is then:

$$\begin{aligned}\phi(x,n,k) &= 1-\prod_{i=1}^{n-k+1}(1-\prod_{\{j:z_{ij}=1\}}x_j)\\ &= 1-\prod_{i=1}^{n-k+1}(1-\prod_{j=i}^{i+k-1}x_j).\end{aligned} \tag{6.23}$$

The structure function $\phi(x,n,k)$, for n=k, is simply $\prod_{j=1}^{k}x_j$, which is the structure function of the well-known series system. For n greater than k, the structure function $\phi(x,n,k)$ satisfied a recursion relationship, as follow:

Lemma 1 (Zakaria (1989))**.** *The structure function $\phi(x,n,k)$ of linear consecutive K-out-of-n:G systems, for $n>k$, satisfies*

$$\phi(x,n,k)=\phi(x,n-1,k)+(1-\phi(x,n-1,k))\prod_{j=n-k+1}^{n}x_j \tag{6.24}$$

Proof. The minimal-path vector representation (6.23) can be written as:

$$\phi(x,n,k)=1-\prod_{i=1}^{n-k}(1-\prod_{j=i}^{i+k-1}x_j)(1-\prod_{j=n-k+1}^{n}x_j) \tag{6.25}$$

$$=1-(1-(1-\prod_{i=1}^{n-k}(1-\prod_{j=i}^{i+k-1}x_j)))(1-\prod_{j=n-k+1}^{n}x_j) \tag{6.26}$$

$$=1-(1-(1-\prod_{i=1}^{n-k}(1-\prod_{j=i}^{i+k-1}x_j)))(1-\prod_{j=n-k+1}^{n}x_j) \tag{6.27}$$

$$=1-(1-\phi(x,n-1,k))(1-\prod_{j=n-k+1}^{n}x_j) \tag{6.28}$$

$$=\phi(x,n-1,k)+(1-\phi(x,n-1,k))\prod_{j=n-k+1}^{n}x_j, \tag{6.29}$$

where equation 6.28 is due to the fact that, according to 6.23 is the structure function of a linear consecutive *k-out-of-(n-1):G* system. □

The reliability of a system is achieve as its structure function ϕ with x replaced by p, under the restriction that ϕ is *"reduced"* form containing no power of any x_i greater than the first.

Theorem 6.3.1 (Zakaria (1989)). *The structure function of a linear consecutive k-out-of-n:G system for $n > k$, can be obtained recursively by using*

$$\begin{aligned}\phi(x,n,k) = \phi(x,n-1,k)+ \\ (1-\phi(x,n-k-1,k))(1-x_{n-k})\prod_{j=n-k+1}^{n} x_j.\end{aligned} \tag{6.30}$$

Proof. Equation 6.24 can also be written as

$$\phi(x,r,k) = \phi(x,r-1,k)(1-\prod_{j=r-k+1}^{r} x_j) + \prod_{j=r-k+1}^{r} x_j. \tag{6.31}$$

Equation 6.30 will be proved by repeated application of equation 6.24 to the second term of recursion formula 6.24.

Substituting 6.31, for *r=n-1*, into 6.24, the result obtain of the first application as follows:

$$\begin{aligned}\phi(x,n,k) &= \phi(x,n-1,k)+ \\ &\quad (1-\phi(x,n-2,k)(1-\prod_{j=n-k}^{n-1} x_j) - \prod_{j=n-k}^{n-1})\prod_{j=n-k+1}^{n} x_j \\ &= \phi(x,n-1,k)+ \\ &\quad (1-\phi(x,n-2,k)(1-x_{n-k}) - x_{n-k})\prod_{j=n-k+1}^{n} x_j \qquad (6.32) \\ &= \phi(x,n-1,k)+ \\ &\quad (1-\phi(x,n-2,k)(1-x_{n-k})\prod_{j=n-k+1}^{n} x_j \qquad (6.33)\end{aligned}$$

where the second equation is due to the fact

$$\prod_{j=n-k}^{n-1} x_j \prod_{j=n-k+1}^{n} x_j = \prod_{j=n-k}^{n} x_j = x_{n-k}\prod_{j=n-k+1}^{n} x_j. \tag{6.34}$$

Substituting 6.31, for *r=n-2*, into the second term of equation 6.34, we obtain the result of the second application as follows:

$$
\begin{aligned}
\phi(x,n,k) &= \phi(x,n-1,k)+ \\
&\quad (1-\phi(x,n-3,k)(1-\prod_{j=n-K-1}^{n-2} x_j)-\prod_{j=n-k-1}^{n-2} x_j) \\
&\quad (1-x_{n-k})\prod_{j=n-k+1}^{n} x_j \qquad (6.35) \\
&= \phi(x,n-1,k)+ \\
&\quad (1-\phi(x,n-3,k)(1-\prod_{j=n-K-1}^{n-k} x_j)-\prod_{j=n-k-1}^{n-k} x_j) \\
&\quad (1-x_{n-k})\prod_{j=n-k+1}^{n} x_j \qquad (6.36) \\
&= \phi(x,n-1,k)+ \\
&\quad (1-\phi(x,n-3,k)(1-\prod_{j=n-K-1}^{n-k} x_j) \\
&\quad (1-x_{n-k})\prod_{j=n-k+1}^{n} x_j \qquad (6.37) \\
&= \phi(x,n-1,k)+ \\
&\quad (1-\phi(x,n-3,k))(1-x_{n-k})\prod_{j=n-k+1}^{n} x_j \qquad (6.38)
\end{aligned}
$$

where the second equation (6.35) is due the fact that

$$\prod_{j=n-k}^{n-2} x_j \prod_{j=n-k+1}^{n} x_j = \prod_{j=n-k-1}^{n} x_j = \prod_{j=n-k-1}^{n-k} x_j \prod_{j=n-k+1}^{n} x_j.$$

and the last equation 6.38 is due to the fact that:

$$(1-\prod_{j=n-k-1}^{n-k} x_j)(1-x_{n-k}) = (1-x_{n-k})$$

From equation 6.33 and 6.38 obtain the result of the $(k-1)^{th}$ application as

$$
\begin{aligned}
\phi(x,n,k) &= \phi(x,n-1,k)+ \\
&\quad (1-\phi(x,n-k,k))(1-x_{n-k})\prod_{j=n-k+1}^{n} x_j \qquad (6.39)
\end{aligned}
$$

Finally, substituting (6.31), for $r = n - k$, into the second term of equation 6.39 result the kth application as follows:

$$\begin{aligned}
\phi(x,n,k) &= \phi(x,n-1,k)+ \\
&\quad (1-\phi(x,n--1k,k)(1-\prod_{j=n-2k-1}^{n-k} x_j) - \prod_{j=n-2k-1}^{n-k} x_j) \\
&\quad (1-x_{n-k}) \prod_{j=n-k+1}^{n} x_j \qquad (6.40) \\
&= \phi(x,n-1,k)+ \\
&\quad (1-\phi(x,n-k-1,k)(1-\prod_{j=n-2K-1}^{n-k} x_j) \\
&\quad (1-x_{n-k}) \prod_{j=n-k+1}^{n} x_j \qquad (6.41) \\
&= \phi(x,n-1,k)+ \\
&\quad (1-\phi(x,n-k-1,k))(1-x_{n-k}) \prod_{j=n-k+1}^{n} x_j \qquad (6.42)
\end{aligned}$$

where the last equation, which is due to the fact that:

$$(1-\prod_{j=n-2k+1}^{n-k} x_j)(1-x_{n-k}) = (1-x_{n-k})$$

is the "reduced" structure function as stated in this theorem

□

Corollary 1 (Zakaria (1989)). *For $k \leq n \leq 2k$, the structure function of a linear consecutive k-out-of-n:G system can be written as*

$$\phi(x,n,k) = \prod_{j=1}^{k} x_j + \sum_{i=1}^{n-k}(1-x_i) \prod_{j=i+1}^{i+k} x_j \qquad (6.43)$$

Proof. When $k \leq n \leq 2k$, the value of the structure function $\phi(x,n-k-1,k)$ is equal to 0, so that, by using (6.30),

$$\phi(x,n,k) = \phi(x,n-1,k) + (1-x_{n-k}) \prod_{j=n-k+1}^{n} x_j \qquad (6.44)$$

Equation (6.43) is then verified recursively as follows:

$$\phi(x,k,k) = \prod_{j=1}^{k} x_j$$
$$\phi(x,k+1,k) = \prod_{j=1}^{k} x_j + (1-x_1)\prod_{j=2}^{k+1} x_j$$
$$\phi(x,k+2,k) = \prod_{j=1}^{k} x_j + (1-x_1)\prod_{j=2}^{k+1} x_j + (1-x_2)\prod_{j=3}^{k+2} x_j$$

Finally

$$\begin{aligned}\phi(x,n,k) &= \prod_{j=1}^{k} x_j + (1-x_1)\prod_{j=2}^{k+1} x_j + (1-x_2)\prod_{j=3}^{k+2} x_j + \\ &\quad \cdots + (1-x_{n-k})\prod_{j=n-k+1}^{n} x_j \\ &= \prod_{j=1}^{k} x_j + \sum_{i=1}^{n-k}(1-x_i)\prod_{j=i+1}^{i+k} x_j\end{aligned}$$

□

Example 6.3.1:
Consider the linear consecutive *2-out-of-5:G system*. This system is working when at least two consecutive components are functioning. Here $k = 2$ and $n = 5$. Using the equation 6.23 the structure function of this system is:

$$\phi(x,5,2) = 1 + \prod_{i=1}^{4}(1 - \prod_{j=i}^{i+1} x_j) \tag{6.45}$$
$$\begin{aligned}&= 1 - (1-x_1x_2)(1-x_2x_3)(1-x_3x_4)(1-x_4x_5) \\ &= x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5 - x_1x_2x_3 - x_2x_3x_4 \\ &\quad - x_3x_4x_5 - x_1x_2x_4x_5 + x_1x_2x_3x_4x_5.\end{aligned} \tag{6.46}$$

This simplification of this function from equation 6.45 to 6.46 becomes very tedious when n is large. Using the recursion relationship 6.24 the structure function of this system became:

$$\phi(x,2,2) = x_1x_2$$

$$\begin{aligned}\phi(x,3,2) &= \phi(x,2,2) + (1-\phi(x,2,2))\prod_{j=2}^{3} x_j \\ &= x_1x_2 + (1-x_1x_2)x_2x_3 \\ &= x_1x_2 + x_2x_3 - x_1x_2x_3. \end{aligned} \tag{6.47}$$

$$\phi(x,4,2) = \phi(x,3,2) + (1-\phi(x,3,2))\prod_{j=3}^{4} x_j \tag{6.48}$$

$$= x_1x_2 + x_2x_3 - x_1x_2x_3 + (1-(x_1x_2 + x_2x_3 - x_1x_2x_3))x_3x_4 \tag{6.49}$$

$$= x_1x_2 + x_2x_3 + x_3x_4 - x_1x_2x_3 - x_2x_3x_4 \tag{6.50}$$

Therefore, the structure function of a linear consecutive *2-out-of-5:G system* is:

$$\begin{aligned}\phi(x,5,2) &= \phi(x,4,2) + (1-\phi(x,4,2))\prod_{j=4}^{5} x_j \\ &= x_1x_2 + x_2x_3 + x_3x_4 - x_1x_2x_3 - x_2x_3x_4 + \\ &\quad (1-(x_1x_2 + x_2x_3 + x_3x_4 - x_1x_2x_3 - x_2x_3x_4))x_4x_5 \end{aligned} \tag{6.51}$$

$$\begin{aligned} &= x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5 - x_1x_2x_3 - x_2x_3x_4 \\ &\quad - x_3x_4x_5 - x_1x_2x_4x_5 + x_1x_2x_3x_4x_5. \end{aligned} \tag{6.52}$$

Using 6.30 and 6.43, the structure function of this system became as follows:

$$\phi(x,2,2) = x_1x_2$$

$$\begin{aligned}\phi(x,3,2) &= \phi(x,2,2) + (1-x_1)\prod_{j=2}^{3} x_j \\ &= x_1x_2 + (1-x_1)x_2x_3 \\ \phi(x,4,2) &= \phi(x,3,2) + (1-x_2)\prod_{j=3}^{4} x_j \\ &= x_1x_2 + (1-x_1)x_2x_3 + (1-x_2)x_3x_4. \end{aligned}$$

Therefore, the structure function of a linear consecutive *2-out-of-5:G system* is:

$$\begin{aligned}\phi(x,5,2) &= \phi(x,4,2) + (1-\phi(x,2,2))(1-x_3)\prod_{j=4}^{5} x_j \\ &= x_1x_2 + (1-x_1)x_2x_3 + (1-x_2)x_3x_4 + (1-x_1x_2)(1-x_3)x_4x_5 \\ &\quad (1 - x_1x_2 + x_2x_3 + x_3x_4 - x_1x_2x_3 - x_2x_3x_4)x_4x_5 \qquad (6.53) \\ &= x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5 - x_1x_2x_3 - x_2x_3x_4 \\ &\quad - x_3x_4x_5 - x_1x_2x_4x_5 + x_1x_2x_3x_4x_5. \qquad (6.54)\end{aligned}$$

6.3.2 Reliability functions of linear consecutive k-out-of-n:G systems

The reliability of a system is the probability that is structure function $\phi(x,n,k)$ equals 1, which, since ϕ is an indicator variable, equals its expectation:

$$R(P,n,k) = P(\phi(x,n,k)=1) = E(\phi(x,n,k)). \qquad (6.55)$$

For a system with independent components, $\mathbf{R}$ may be found, simply, by replacing $\mathbf{x}$ by $\mathbf{p}$ in the "reduced" structure function ϕ.

The reliability function of linear consecutive *k-out-of-n:G* systems when all components are independent and $n \geq k$ is obtainable recursively by:

$$R_G(p,n,k) = R_G(p,n-1,k) + (1 - R_G(p,n-k-1,k))q_{n-k}\prod_{j=n-k+1}^{n} p_j \qquad (6.56)$$

if $p_1 = p_2 = \cdots = p_n = p$, the reliability function becomes

$$R_G(p,n,k) = R_G(p,n-1,k) + (1 - R_G(p,n-k-1,k))q_{n-k}p^k \qquad (6.57)$$

if $k \leq n \leq 2k$, the reliability function is obtainable by

$$R_G(p,n,k) = \prod_{j=1}^{k} p_j + \sum_{i=1}^{n-k} q_i \prod_{j=i+1}^{i+k} p_j \qquad (6.58)$$

if $k \leq n \leq 2k$ and $p_1 = p_2 = \cdots = p_n = p$ the reliability function is obtainable by

$$R_G(p,n,k) = ((n-k)(1-p)+1)p^k. \qquad (6.59)$$

To compute system reliability, especially for a large system, equation 6.56 can be used directly to produce algorithm. The algorithm should begin with reading and checking the input n, k and p_j such that $1 \leq k \leq n$ and $0 \leq p_j \leq 1$. The next step is to compute $R_G(p,k,k) = \prod_{j=1}^{k} p_j$, and the last step is to compute $R_G(p,n,k)$ using 6.56 or using 6.58 for a system with $k \leq n \leq 2k$.

6.3.3 Structure of linear consecutive k-out-of-n:F systems

The set of minimal-cut vectors of linear consecutive K-out-of-n:F systems is:

$$W = \{W_i\}_{i=1}^{n-k+1} = \{(0_k, 1_{n-k}), (1, 0_k, 1_{n-k-1}), ..., (1_{n-k}, 0_k)\}. \tag{6.60}$$

where 0_j is the j-dimensional zero vector (0,0,...,0) and 1_j is the j-dimensional unit vector (1,1,...,1). There are $n-k+1$ minimal-cut vectors since there are $n-k+1$ possibilities of placing the k consecutive zeroes out of n locations. Using the minimal-cut vector representation the structure function for the linear consecutive K-out-of-n:F systems is then

$$\begin{aligned}\psi(x,n,k) &= 1 - \prod_{i=1}^{n-k+1}(1 - \prod_{\{j:w_{ij}=0\}}(1-x_j)) \\ &= \prod_{i=1}^{n-k+1}(1 - \prod_{j=i}^{i+k-1}(1-x_j)).\end{aligned} \tag{6.61}$$

The structure function of this system $\psi(x,n,k)$ can also be represented in a recursion relationship, as shown in the next lemma:

Lemma 2 (Zakaria (1989)). *The structure function $\psi(x,n,k)$ of linear consecutive K-out-of-n:F systems, for $n > k$, can be expressed as*

$$\psi(x,n,k) = \psi(x,n-1,k) - \psi(x,n-1,k)) \prod_{j=n-k+1}^{n}(1-x_j) \tag{6.62}$$

Proof. The minimal-cut vector representation can be written as

$$\psi(x,n,k) = \prod_{i=1}^{n-k}(1 - \prod_{j=i}^{i+k-1}(1-x_j))(1 - \prod_{j=n-k+1}^{n}(1-x_j)) \tag{6.63}$$

$$= \psi(x,n-1,k))(1 - \prod_{j=n-k+1}^{n}(1-x_j)) \tag{6.64}$$

$$= \psi(x,n-1,k) - \psi(x,n-1,k) \prod_{j=n-k+1}^{n}(1-x_j), \tag{6.65}$$

□

As in the *"G"* systems, this recursion doesn't directly produce a "reduced" form of structure function. The following theorem gives a recursion relationship that directly produce a structure function of linear consecutive *K-out-of-n:F* system in "reduced" form.

Theorem 6.3.2 (Zakaria (1989)). *The structure function of a linear consecutive k-out-of-n:F system for $n > k$, can be obtained recursively by using*

$$\psi(x,n,k) = \psi(x,n-1,k) + \psi(x,n-k-1,k)x_{n-k}\prod_{j=n-k+1}^{n}(1-x_j). \tag{6.66}$$

Proof. A linear consecutive k-out-of-n:F system can be regarded as the dual of consecutive *k-out-of-n:G* system since $w_i = 1 - z_i$. Using the duality definition of *Barlow and Proschan (1981)*, the structure function of a linear consecutive *k-out-of-n:F* system can be write as:

$$\psi(x,n,k) = 1 - \phi(1-x,n,k). \tag{6.67}$$

where $\phi(1-x,n,k)$ is the structure function of a linear consecutive *k-out-of-n:F* in the context of failure. Replacing x by *1-x* in 6.30, and using 6.67, it's result:

$$\begin{aligned}\psi(x,n,k) &= (1-\phi(1-x,n-1,k)) - \\ &\quad (1-\phi(1-x,n-k-1,k)x_{n-k}\prod_{j=n-k+1}^{n}(1-x_j) \\ &= \psi(x,n-1,k) - \\ &\quad \psi(x,n-k-1,k)x_{n-k}\prod_{j=n-k+1}^{n}(1-x_j)\end{aligned} \tag{6.68}$$

which verifies equation 6.66

□

Corollary 2 (Zakaria (1989)). *For $k \le n \le 2k$, the structure function of a linear consecutive k-out-of-n:F system can be written as*

$$\psi(x,n,k) = 1 - \prod_{j=1}^{k}(1-x_j) - \sum_{i=1}^{n-k}(1-x_i)\prod_{j=i+1}^{i+k}(1-x_j). \tag{6.69}$$

Proof. When $k \le n \le 2k$, the value of the structure function $\psi(x,n-k-1,k)$ is equal to 1, so that, by using (6.66),

$$\psi(x,n,k) = \psi(x,n-1,k) + x_{n-k})\prod_{j=n-k+1}^{n}(1-x_j). \tag{6.70}$$

Equation (6.69) is then obtainable and hence verified using the similar procedure employed in corollary 1 □

Example 6.3.2:
Consider the linear consecutive *2-out-of-5:F system*. This system is failed when at least two consecutive components are failed. Here $k = 2$ and $n = 5$. Using the equation 6.61 we can write the structure function of this system as

$$\begin{aligned}
\psi(x,5,2) &= \prod_{i=1}^{4}(1-\prod_{j=i}^{i+1}(1-x_j)) \\
&= (1-(1-x_1)(1-x_2))(1-(1-x_2)(1-x_3) \\
&\quad (1-(1-x_3)(1-x_4))(1-(1-x_4)(1-x_5) \\
&= x_2x_4 + x_1x_3x_4 + x_1x_3x_5 + x_2x_3x_5 - x_1x_2x_3x_4 - \\
&\quad x_1x_3x_4x_5 - x_1x_2x_3x_5 - x_2x_3x_4x_5 - x_1x_2x_3x_4x_5.
\end{aligned}$$

Using the recursion relationship 6.62, we can write the structure function of this system as

$$\begin{aligned}
\psi(x,2,2) &= x_1 + x_2 - x_1x_2 \\
\psi(x,3,2) &= \psi(x,2,2) - \psi(x,2,2)(1-x_2)(1-x_3) \\
&= x_2 + x_1x_3 - x_1x_2x_3 \\
\psi(x,4,2) &= \psi(x,3,2) - \psi(x,3,2)(1-x_3)(1-x_4) \\
&= x_1x_3 + x_2x_3 + x_2x_4 - x_1x_2x_3 - x_2x_3x_4 \\
\psi(x,5,2) &= \psi(x,4,2) - \psi(x,4,2)(1-x_4)(1-x_5) \\
&= x_2x_4 + x_1x_3x_4 + x_1x_3x_5 + x_2x_3x_5 - x_1x_2x_3x_4 - \\
&\quad x_1x_3x_4x_5 - x_1x_2x_3x_5 - x_2x_3x_4x_5 - x_2x_3x_4x_5 - x_1x_2x_3x_4x_5.
\end{aligned} \tag{6.71}$$

Using 6.66, we obtain:

$$
\begin{aligned}
\psi(x,2,2) &= x_1 + x_2 - x_1x_2 \\
\psi(x,3,2) &= \psi(x,2,2) - x_1(1-x_2)(1-x_3) \\
&= x_2 + x_1x_3 - x_1x_2x_3 \\
\psi(x,4,2) &= \psi(x,3,2) + x_2(1-x_3)(1-x_4) \\
&= x_1x_3 + x_2x_3 + x_2x_4 - x_1x_2x_3 - x_2x_3x_4 \\
\psi(x,5,2) &= \psi(x,4,2) - x_3(1-x_4)(1-x_5) \\
&= x_2x_4 + x_1x_3x_4 + x_1x_3x_5 + x_2x_3x_5 - x_1x_2x_3x_4 - \\
&\quad x_1x_3x_4x_5 - x_1x_2x_3x_5 - x_2x_3x_4x_5 - x_2x_3x_4x_5 - x_1x_2x_3x_4x_5
\end{aligned} \tag{6.72}
$$

6.3.4 Reliability functions of linear consecutive k-out-of-n:F systems

The reliability functions of linear consecutive *K-out-of-n:F* systems are stated as follows:

The reliability function of linear consecutive *K-out-of-n:F* systems when all components are independent and $n \geq k$ is obtainable recursively through:

$$R_F(p,n,k) = R_F(p,n-1,k) - R_F(p,n-k-1,k)p_{n-k} \prod_{j=n-k+1}^{n} q_j \tag{6.73}$$

if $p_1 = p_2 = \cdots = p_n = p$, the reliability function becomes

$$R_F(p,n,k) = R_F(p,n-1,k) - R_F(p,n-k-1,k)pq^k \tag{6.74}$$

if $k \leq n \leq 2k$, the reliability function is

$$R_F(p,n,k) = 1 - \prod_{j=1}^{k} q_j - \sum_{i=1}^{n-k} p_i \prod_{j=i+1}^{i+k} q_j \tag{6.75}$$

if $k \leq n \leq 2k$ and $p_1 = p_2 = \cdots = p_n = p$ the reliability function is

$$R_F(p,n,k) = 1 - ((n-k)p+1)q^k. \tag{6.76}$$

Using 6.73 and 6.75, an algorithm can be produced to compute the reliability of this system. This can easily be done by modifying the algorithm for the "G" systems. In this case, use input q_j instead of p_j to obtain system unreliability. Then subtract this quantity from 1 to obtain system reliability.

6.3.5 Simulation of linear consecutive k-out-of-n

To compute a simulation of Monte Carlo linear consecutive *k-out-of-n*, using Weibull distribution the follow algorithm is developed:

Step 1: Define the function of the structure of linear consecutive *k-out-of-n* system

Step 2: Calculate the time to censoring - t_c with the parameters of distribution choose

Step 3: Generate t_i from random distribution function

Step 4: Compare the time t_i with T_c to each component and give $x_i = 0$ if are above or $x_i = 1$ below the t_c

Step 5: With x_i and structure function calculate if the system are working or not

Step 6: Repeat for M times (the dimension of the cycle simulation)

Step 7: Calculate the reliability: the number of times that the system are working for the number of samples M

The program has been written in *R* software, and in the beginning it's define the function structure and simulations and then applied the loop *"for"* to made the cycle and the reliability calculation. The difference between linear consecutive *k-out-of-n:F* system and the linear consecutive *k-out-of-n system:G* is the structure function of the system, that in the program is define by a function with the name *str fun*.

```
bet=beta[j]
tcm=0;fiab=0;tcm1=0
for (k in 1:4){
    cen <- ceni[k]
    tc <- scx1*(-log(cen))^(1/shx1)
    for (i in 1:length(mi)){
    m=mi[i]
    rest=simul_fun(m)
    fiab[i]=rest/m
    }
    tcm=cbind(tcm,fiab)
    cens=censtr[k]
    bi=bstr[j]
}
tcm1=tcm[,-1]
bi=bstr[j]
```

Algorithm 11: Simulation of linear consecutive k-out-of-n

6.3.6 Simulation of linear consecutive 2-out-of-5:G

The simulation of linear consecutive *2-out-of-5:G* to all components used the same Weibull parameters; the shape parameter β have the values *0,5;1;1,5 and 2*; the scale parameter is $\eta = 10$, for all components and simulations. The simulation is made for different number

of samples to verified the impact of the number of samples for each simulation. Another interest characteristic is to simulate reliability with different censored data, in this case it's used 5%,10%, 20% and 30%.

	$\mathbf{C}_{5\%}$				$\mathbf{C}_{10\%}$				$\mathbf{C}_{20\%}$				$\mathbf{C}_{30\%}$			
Sample	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2
10	1	1	0.9	0.7	1	1	0.9	0.8	1	1	0.9	0.9	1	1	1	1
100	0.95	0.95	0.83	0.83	0.99	0.99	0.93	0.86	1	1	0.97	0.87	1	1	0.98	0.88
500	0.96	0.91	0.9	0.81	1	0.98	0.94	0.88	1	1	0.97	0.9	1	1	0.99	0.92
1000	0.96	0.92	0.89	0.84	1	0.99	0.94	0.85	1	1	0.98	0.88	1	1	1	0.92
2000	0.95	0.93	0.88	0.82	1	0.99	0.94	0.87	1	1	0.98	0.9	1	1	1	0.92

Table 6.3: Simulation 2-out-of-5:G, Weibull $(\beta, \%C, n)$, $\eta = 10$ $M = 100$

The results are explicit in the table 6.3. From the analysis of the table, it's possible to see that with the increase of the sample number the value of reliability stabilized at a certain value. With the increase of the shape factor β, the value of reliability decreases, not so much, but can appoint the exponential shape as the most favourable state to have a higher value of reliability. The reliability decreases smoothly with the increase of censored data. One possible explanation is that with the increase of censorship, the system is not so stable and the faults are a little more and have more impact on reliability.

Test with other parameters and also make a comparative test with different parameters for each component will be recommend.

6.3.7 Simulation of linear consecutive 2-out-of-5:F

The simulation of linear consecutive *2-out-of-5:F* are similar with *2-out-of-5:G* and used the same Weibull parameters; the shape parameter β have the values *0,5;1;1,5 and 2* and the the scale parameter is $\eta = 10$. The simulation is made for different number of samples to verified the impact of the number of samples for each simulation.

	$\mathbf{C}_{5\%}$				$\mathbf{C}_{10\%}$				$\mathbf{C}_{20\%}$				$\mathbf{C}_{30\%}$			
Sample	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2	$\beta_{0.5}$	β_1	$\beta_{1.5}$	β_2
10	1	0.8	0.7	0.5	1	1	0.7	0.5	1	1	1	0.8	1	1	1	1
100	0.89	0.8	0.79	0.63	0.98	0.98	0.86	0.71	1	0.99	0.97	0.78	1	1	1	0.81
500	0.91	0.87	0.74	0.65	0.98	0.98	0.87	0.71	1	1	0.95	0.77	1	1	0.99	0.8
1000	0.88	0.85	0.76	0.67	0.99	0.96	0.86	0.74	1	1	0.94	0.77	1	1	0.98	0.83
2000	0.9	0.84	0.75	0.68	0.99	0.96	0.88	0.7	1	0.99	0.94	0.79	1	1	0.98	0.82

Table 6.4: Simulations 2-out-of-5:F, Weibull $(\beta, \%C, n)$, $\eta = 10$

The results are explicit in the table 6.4 and compare with the simulation of reliability *2-out-of-5:G* system the evolution and the pattern are the same and the absolute values of reliability for each simulation is very close.

Test with other parameters and also make a comparative test with different parameters for each component will be recommend.

6.4 Importance components

6.4.1 Structural importance of components

In a system some components are more important for the system reliability than other components.

The importance measure may be used to arrange the components in order of increasing or decreasing importance. In the process system functions, a component may be very important for the essential safety function but may have little importance for the other system functions.

A components in series in the system is a cut set of order *1* and is generally more important than a component in parallel or member of a cut set of higher order.

The reliability of an equipment or system may be improved by introducing redundant component or by using a higher quality component or another operational parameter like production and environmental conditions or type of maintenance.

The objective of the component importance measure is to help to allocate inspection and maintenance resources to the most important components or improve better preventive maintenance tasks.

6.4.2 Birnbaum's measure

Birnbaum (1969) proposed the following measure of importance of component i at time t is:

$$I^B(i|t) = \frac{\partial h(p(t))}{\partial p_i(t)} \quad for \quad i = 1, 2, ..., n. \tag{6.77}$$

And can be expresses as:

$$I^B(i|t) = \frac{\partial h(p(t))}{\partial p_i(t)} = h(1_i, p(t)) - h(0_i, p(t)) \tag{6.78}$$

Remember that $h(\cdot_i, p(t)) = E[\phi(\cdot_i, X(t))]$, such that 6.78 can be written:

$$\begin{aligned} I^B(i|t) &= E[\phi(1_i, X(t))] - E[\phi(0_i, X(t))] \\ &= E[\phi(1_i, X(t)) - \phi(0_i, X(t))] \end{aligned}$$

When $\phi(X(t))$ is a coherent structure, $\phi(1_i, X(t)) - \phi(0_i, X(t))$ can only take on the values 0 and 1. Birnbaums measure can therefore be written as

$$I^B(i|t) = Pr(\phi(1_i, X(t)) - \phi(0_i, X(t)) = 1) \tag{6.79}$$

This is to say that $I^B(i|t)$ is equal to the probability that $(1_i, X(t))$ is a critical path vector for component i at time t.

6.4.3 Improvement potential

When component i is replaced by a perfect component, that is, a component such that $p_i(t) = 1$, the difference between $h(1_i, p(t))$ and *h(p(t))* is called the improvement potential *(IP)* with respect to component i and denoted by $I^{IP}(i|t)$.

The improvement potential with respect to component i at time t is

$$I^{IP}(i|t) = h(1_i, p(t)) - h(p(t)) \quad for \quad i = 1, 2, ..., n. \tag{6.80}$$

or, simplify with Birnbaum measure

$$I^{IP}(i|t) = I^B(i|t)\dot{(}1 - p(t)) \tag{6.81}$$

or, by using the fault tree notation

$$I^{IP}(i|t) = I^B(i|t)q_i(t) \tag{6.82}$$

6.4.4 Risk Achievement Worth

The Risk Achievement Worth (RAW) has been introduced as a risk importance measure in probabilistic safety assessments of nuclear power stations *EPRI (1995)*.

The importance measure Risk Achievement Worth with respect to component i at time t is

$$I^{RAW}(i|t) = \frac{1 - h(0_i, p(t))}{1 - h(p(t))} \quad for \quad i = 1, 2, ..., n. \tag{6.83}$$

The RAW presents a measure of the importance (*worth*) of component i in the actual system reliability and is the ratio of the(conditional) system unreliability if component i is always failed with the actual system unreliability.

6.4.5 Risk Reduction Worth

The Risk Reduction Worth (RRW), $I^{RRW}(i|t)$, is the ratio of the actual system unreliability with the (conditional) system unreliability if component i is replaced by a perfect component with $p_i(t) \equiv 1$.

The importance measure Risk reduction worth (RRW) with respect to component i at time t is

$$I^{RRW}(i|t) = \frac{1 - h(p(t))}{1 - h(1_i, p(t))} \quad for \quad i = 1, 2, ..., n. \tag{6.84}$$

If $I^{RRW}(i|t) \geq 1$ it's possible to write:

$$I^{RRW}(i|t) = \left(1 - \frac{I^{IP}(i|t)}{1 - h(p(t))}\right)^{-1}$$

By fault tree notation and 6.82 RRW stated as follows:

$$I^{RRW}(i|t) = \left(1 - \frac{I^{IP}(i|t)}{Q_0(t)}\right)^{-1} = \left(1 - \frac{I^{B}(i|t)\dot{q}_i(t)}{Q_0(t)}\right)^{-1} \tag{6.85}$$

6.4.6 Criticality Importance

The criticality importance $I^{CR}(i|t)$ of component i at time t is the probability that component i is failed at time t and the system is failed at time t.

$$I^{CR}(i|t) = \frac{I^{B}(i|t) \cdot (1 - p_i(t))}{1 - h(p(t))} \tag{6.86}$$

By using the fault tree notation, $I^{CR}(i|t)$ may be written

$$I^{CR}(i|t) = \frac{I^{B}(i|t) \cdot (1 - q_i(t))}{Q_0(t)} \tag{6.87}$$

Criticality importance criteria is related to Birnbaum's measure and is particularly suitable for prioritizing maintenance actions.

From equation 6.85 the Risk Reduction Worth (RRW) can be expressed as a function of criticality importance (CI):

$$I^{RRW}(i|t) = (1 - I^{CR}(i|t))^{-1} = \frac{1}{1 - I^{CR}(i|t)} \tag{6.88}$$

6.4.7 Simulation Importance components

To program a simulation of Monte Carlo importance components to coherent systems and linear consecutive *k-out-of-n:F and G*, using Weibull distribution the following algorithm is developed:

Step 1: Define the parameters of distribution and initialize the variables to use

Step 2: Define the function of the structure of the system

Step 3: Generate U_i random uniform *(0,1)*

Step 4: Generate t_i from step (1) and step (4), or directly from random distribution function

Step 5: Generate de x_i from the probability of t_i for each component to Weibull distribution

Step 6: With x_i and structure function calculate the importance components

Step 7: Repeat step (3) to step (6) for M times (the dimension of the simulation cycles)

Step 8: Calculate the mean of importance component

The program has been written in R software, and in the beginning it's define the function structure and simulations and then applied the loop *"for"* to made the cycle and calculate the indicators of importance components. The difference from: 1. coherent system; 2. linear consecutive *k-out-of-n:F* system and 3. linear consecutive *k-out-of-n system:G* is the structure function of the system, that in the program is defined by a function with the name *str fun*.

```
shx5=1;scx5=10
bet=1;

IB1i=0;IB2i=0;IB3i=0;IB4i=0;IB5i=0
IP1i=0;IP2i=0;IP3i=0;IP4i=0;IP5i=0
RW1i=0;RW2i=0;RW3i=0;RW4i=0;RW5i=0
IC1i=0;IC2i=0;IC3i=0;IC4i=0;IC5i=0
IBMN=0;IPMN= 0;RWMN=0;ICMN=0
mi <- c(10,100,500,1000,2000)


for (j in 1:length(mi)){

m=mi[j]

for(i in 1:m) {

#simul_fun<-function(){
  t1 <- rweibull(1,bet,scx1)
  t2 <- rweibull(1,bet,scx2)
  t3 <- rweibull(1,bet,scx3)
```

Algorithm 12: Program for simulation importance components (partial)

6.4.8 Simulation complex system: 2-out-of-5 - bridge structure

In this simulation all component have the same Weibull parameters. The shape parameter $\beta = 1$ and the scale parameter is $\eta = 10$, for all components.

The structure function 2-out-of-5 - *bridge structure* may be written:

$$\begin{aligned}\phi(x) = {} & x_1x_4 + x_2x_5 + x_1x_3x_5 + x_2x_3x_4 - x_1x_3x_4x_5 - x_1x_2x_3x_5 \\ & - x_1x_2x_3x_4 - x_2x_3x_4x_5 - x_1x_{2x4}x_5 + 2x_1x_2x_3x_4x_5 \end{aligned} \tag{6.89}$$

From equation 6.89 it's possible to achieve and measure the importance components beginning to express the Birnbaum's measure $I^B(i|t)$ to all components:

$$\begin{aligned} IB1 = \frac{\partial\phi(x)}{\partial x_1} = {} & x_4 + 0 + x_3x_5 + 0 - x_3x_4x_5 - x_2x_3x_5 \\ & - x_2x_3x_4 - 0 - x_2x_4x_5 - 2x_2x_3x_4x_5 \end{aligned}$$

$$IB2 = \frac{\partial \phi(x)}{\partial x_2} = 0 + x_5 + 0 + x_3x_4 - 0 - x_1x_3x_5 - x_1x_3x_4 - x_3x_4x_5 - x_1x_4x_5 + 2x_1x_3x_4x_5$$

$$IB3 = \frac{\partial \phi(x)}{\partial x_3} = 0 + 0 + x_1x_5 + x_2x_4 - x_1x_4x_5 - x_1x_2x_5 - x_1x_2x_4 - x_2x_4x_5 - 0 + 2x_1x_2x_4x_5$$

$$IB4 = \frac{\partial \phi(x)}{\partial x_4} = x_1 + 0 + 0 + x_2x_3 - x_1x_3x_5 - 0 - x_1x_2x_3 - x_2x_3x_5 - x_1x_2x_5 + 2x_1x_2x_3x_5$$

$$IB5 = \frac{\partial \phi(x)}{\partial x_5} = 0 + x_2 + x_1x_3 + 0 - x_1x_3x_4 - x_1x_2x_3 - 0 - x_2x_3x_4 - x_1x_2x_4 + 2x_1x_2x_3x_4$$

	Import.Birbaum					**Improvement Potential**					**Risk redution Worth**					**Criticality Import.**				
Sample	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
10	0.46	0.43	0.12	0.27	0.38	0.24	0.21	0.05	0.12	0.14	2.3	2	1.2	1.5	2.1	0.44	0.68	0.12	0.23	0.7
100	0.37	0.37	0.12	0.39	0.37	0.18	0.18	0.06	0.18	0.18	2	2	1.2	2.1	2.1	0.37	0.61	0.14	0.36	0.55
500	0.37	0.38	0.13	0.36	0.38	0.18	0.18	0.06	0.18	0.18	2	2.3	1.2	2	2.1	0.4	0.68	0.15	0.34	0.55
1000	0.37	0.37	0.13	0.37	0.38	0.19	0.18	0.06	0.19	0.19	2.1	2.1	1.2	2.2	2.3	0.39	0.61	0.15	0.34	0.52
2000	0.37	0.38	0.12	0.38	0.37	0.19	0.19	0.06	0.19	0.18	2.2	2.2	1.2	2.2	2	0.4	0.62	0.15	0.35	0.49

Table 6.5: Simulation Component Importance, 2-out-of-5, Weib. (n), $\beta = 1$, $\eta = 10$

The results are explicit in the table 6.5. From the analysis of the table the increment of the sample number stabilize the criticality of each component for different criteria. The results show that the components *C3* and *C2* (lower values) are the most importance component to the various criteria. The values of reliability is low from all situations and this is an important information to know the need to do more study or implement some actions to increment the reliability of components and the system.

6.4.9 Simulation Linear consecutive 2-out-of-5:G

The structure function of a linear consecutive *2-out-of-5:G system* is:

$$\phi(x,5,2) = x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5 - x_1x_2x_3 - x_2x_3x_4 \\ - x_3x_4x_5 - x_1x_2x_4x_5 + x_1x_2x_3x_4x_5. \tag{6.90}$$

From equation 6.90 it's possible to achieve and measure the importance components beginning to express the Birnbaum's measure $I^B(i|t)$ to all components::

$$IB1 = \frac{\partial\phi(x)}{\partial x_1} = x_2 + 0 + 0 + 0 - x_2x_3 - 0 - 0 \\ - x_2x_4x_5 + x_2x_3x_4x_5$$

$$IB2 = \frac{\partial\phi(x)}{\partial x_2} = x_1 + x_3 + 0 + 0 - x_1x_3 - x_3x_4 \\ - 0 - x_1x_4x_5 + x_1x_3x_4x_5$$

$$IB3 = \frac{\partial\phi(x)}{\partial x_3} = 0 + x_2 + x_4 + 0 - x_1x_2 - x_2x_4 \\ - x_4x_5 - 0 + x_1x_2x_4x_5$$

$$IB4 = \frac{\partial\phi(x)}{\partial x_4} = 0 + 0 + x_3 + x_5 - 0 - x_2x_3 \\ - x_3x_5 - x_1x_2x_5 + x_1x_2x_3x_5$$

$$IB5 = \frac{\partial\phi(x)}{\partial x_5} = 0 + 0 + 0 + x_4 - 0 - 0 - x_3x_4 \\ - x_1x_2x_4 + x_1x_2x_3x_4$$

	Import.Birbaum					**Improvement Potential**					**Risk redution Worth**					**Criticality Import.**				
Sample	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
10	0.17	0.48	0.46	0.37	0.21	0.08	0.21	0.2	0.18	0.1	1.2	2.9	2.1	2.2	1.3	0.28	0.82	0.85	0.47	0.25
100	0.19	0.43	0.32	0.4	0.19	0.08	0.18	0.13	0.18	0.09	1.5	2.7	1.8	2.7	1.4	0.28	0.74	0.85	0.62	0.3
500	0.18	0.42	0.32	0.43	0.2	0.08	0.19	0.13	0.19	0.08	1.3	2.3	1.8	2.6	1.4	0.26	0.57	0.64	0.79	0.36
1000	0.19	0.43	0.32	0.4	0.19	0.09	0.19	0.14	0.18	0.08	1.4	2.7	1.7	2.5	1.4	0.27	0.6	0.72	0.79	0.37
2000	0.18	0.44	0.31	0.41	0.19	0.08	0.2	0.13	0.19	0.08	1.3	2.7	1.6	2.5	1.4	0.27	0.6	0.66	0.75	0.36

Table 6.6: Simulation Component Importance, 2-out-of-5:G, Weib. (n), $\beta = 1$, $\eta = 10$

The results are explicit in the table (6.6). From the analysis of the table can see that when increment the number of sample the values of importance components stabilize for each component. The components *C1* and *C5* (lower values) are the "critical components", that give the most important for the majority of criteria. This results is understood if looking at the structure of the function and understand that they are the beginner and the final components of the series system.

6.4.9.1 Simulation Linear consecutive 2-out-of-5:F

The structure function of system Linear consecutive 2-out-of-5:F can be written:

$$\psi(x,5,2) = x_2x_4 + x_1x_3x_4 + x_1x_3x_5 + x_2x_3x_5 - x_1x_2x_3x_4 - \\ x_1x_3x_4x_5 - x_1x_2x_3x_5 - x_2x_3x_4x_5 - x_2x_3x_4x_5 - x_1x_2x_3x_4x_5. \tag{6.91}$$

From equation 6.91 it's possible to achieve and measure the importance components beginning to express the Birnbaum's measure $I^B(i|t)$ to all components::

$$IB1 = \frac{\partial\phi(x)}{\partial x_1} = 0 + x_3x_4 + x_3x_5 + 0 - x_2x_3x_4 \\ - x_3x_4x_5 - x_2x_3x_5 - 0 - x_2x_3x_4x_5$$

$$IB2 = \frac{\partial\phi(x)}{\partial x_2} = x_4 + 0 + 0 + x_3x_5 - x_1x_3x_4 - 0 \\ - x_1x_3x_5 - x_3x_4x_5 - x_1x_3x_4x_5$$

$$IB3 = \frac{\partial \phi(x)}{\partial x_3} = 0 + x_1x_4 + x_1x_5 + x_2x_5 - x_1x_2x_4$$
$$- x_1x_4x_5 - x_1x_2x_5 - x_2x_4x_5 - x_1x_2x_4x_5$$

$$IB4 = \frac{\partial \phi(x)}{\partial x_4} = x_2 + x_1x_3 + 0 + 0 - x_1x_2x_3 - x_1x_3x_5$$
$$- 0 - x_2x_3x_5 - x_1x_2x_3x_5$$

$$IB5 = \frac{\partial \phi(x)}{\partial x_5} = 0 + 0 + x_1x_3 + x_2x_3 - 0 - x_1x_3x_4$$
$$- x_1x_2x_3 - x_2x_3x_4 - x_1x_2x_3x_4$$

	Import.Birbaum					**Improvement Potential**					**Risk redution Worth**					**Criticality Import.**				
Sample	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
10	-0.08	0.28	0.14	0.36	0.12	-0.03	0.12	0.08	0.21	0.08	0.96	2	1.5	2.3	1.3	-0.16	0.68	0.12	0.64	0.1
100	0.05	0.29	0.17	0.31	0.06	0.04	0.18	0.13	0.19	0.05	1.2	1.7	1.4	2.1	1.2	0.03	0.31	0.1	0.27	0.05
500	0.07	0.31	0.19	0.33	0.08	0.06	0.2	0.13	0.21	0.06	1.2	2.1	1.4	2	1.2	0.01	0.38	0.13	0.29	0.06
1000	0.05	0.32	0.18	0.31	0.06	0.04	0.2	0.13	0.2	0.05	1.1	2.2	1.4	2.1	1.2	0	0.41	0.11	0.32	0.05
2000	0.06	0.3	0.19	0.32	0.06	0.05	0.19	0.13	0.2	0.05	1.2	2.3	1.4	2.2	1.2	0.01	0.39	0.11	0.33	0.04

Table 6.7: Simulation Component Importance, 2-out-of-5:F, Weib. (n), $\beta = 1$, $\eta = 10$

The results are explicit in the table 6.7. From the analysis of the table the components*C1* (lowest value) is the most important for the various criteria, followed by components *C3* and *C5*. The explanation can be obtained by analysing the structure of the function and understanding that it is the initial component that is most present in the structural function.

Test with other parameters and also make a comparative test with different parameters for each component will be recommend.

Chapter 7

Reliability analysis and simulation - case study

Learning about the existing maintenance procedures for equipment and the statistical analysis of the field data is an essential step toward developing an optimal maintenance plan. Maintenance defines the set of activities performed on an item to retain it in or to restore it to a specific state.

Unexpected failures usually have adverse effects and may result in major accidents. There is a strong relationship between maintenance practices and the occurrence of major accidents. Productivity is closely related to the availability and reliability of the equipment. The major challenge for a maintenance engineer is to implement a maintenance strategy, which maximizes the availability and efficiency of the equipment, controls the rate of equipment deterioration, ensures its safe operation, and minimizes the total cost of the operation. This can only be achieved by adopting a structured approach to the study of equipment degradation and failure and designing an optimum strategy for inspection and maintenance.

Due to the increasing complexity of modern control systems and the growing demand for quality, cost eciency, availability, reliability, and safety, the analysis of failure systems in complex industries is gaining more and more importance.

Centrifugal pumps are widely used in the petrochemical industry and, in some instances, the number of pumps used could easily amount to hundreds of pumps in a typical petrochemical plant. Consequently, the reliability of these pumps essentially translates into stable and reliable plant operation as the pumps performances are critical to ensure continuous plant productivity. Repairable systems, such as centrifugal pumps, consist of a large number of interacting components that perform the system's required functions.

This chapter describes a methodology that allows identifying the most critical components of the system and also helps as a tool to analyse maintenance performance work. The aim is to obtain the best inspection intervals and to optimize cost, reliability and risk with the use of simulation and calculation tools. This approach also intends to contribute to the

fact that the classical reliability analysis performed in industrial companies can evolve to adopt new models of reliability, as in the case of the consecutive *k-out-of-n* linear models applied in the petrochemical industry.

The approach is original and enables, above all, showing the hypotheses of using it in an operational context and taking it into account to define the maintenance policy for critical company equipment.

The chapter begins by describing the equipment that is object of study and its connections, in this case, the centrifugal pump in the petrochemical industry and its components. Next, it explains the methodology proposed, its schematic flowchart, as well as some of the most relevant aspects. In this case, the centrifugal pump and its components are decomposed and regrouped, as necessary and taking into account the maintenance practices of the company under study, the international standards, and other work done in this area, in order to reduce the model to a simple serial system.

The data analysis, from exhaustive fieldwork, will allow us to extract data that serves to illustrate and exemplify our methodology. This section explains how the data was collected and organized, as well as the encountered difficulties and problems. It should be pointed out that our objective was not to make a rigorous estimation of reliability parameters, but rather to obtain data in order to mark and scale the component systems in order to feed our models and simulate and validate our methodology.

It should be noted that the systems were considered in the adult phase of life and, consequently, with a constant average failure rate.

In this chapter the data was analysed and are applied it in the models simulated in the previous chapters and the components are classified by degrees of importance for three types of intervals: semester, semi-annually and annually. The scenarios are compared, the results criticized and finally, various conclusions are reached.

7.1 Information systems and methodological framework

Information in modern society has become so important that we can say that are living in the information age. All human life revolves around information that can be used for the most diverse purposes: to generate knowledge, value, wealth, power. In organizations, information is ubiquitous and has become a critical success factor.

In this sense, information management is crucial for both individuals and organizations. Computational capacity have brought a revolution in this field, being, nowadays, not only a tool for management assistance, but also, in many cases, an integral part of the organizations operational process.

The enormous amount of information that must be managed in maintenance field is, therefore, of vital importance for a better knowledge of equipment behaviour during its life-cycle. In most cases introducing technical modifications with the aim of eliminating breakdowns or reducing their consequences. Quick changes within the market, technology,

etc., have created tensions in the existing organizational forms. Traditional structures are highly bureaucratic and experience shows that they cannot respond quickly enough to an ever-changing environment. Thus, the use of simulation tools allows top management or other management structure to be highly dynamic and to respond quickly to situations developed within or outside the company.

The function of a computer is to help people perform their work. There will certainly be many and diverse views as to the value and impact of information systems and simulation in people's lives, which clearly makes how they are developed and used, critical.

Conceptual modelling and simulation is fundamental for creating a model that is as close as possible to a real system. Without a correct definition of the whole system, from the representation of the intervening entities and processes to the definition of the relations between them, it is not possible to even consider the possibility of designing a reliable model.

Companies, and especially maintenance managers, are very open minded to jobs of this nature that help them think about the organization and indicate valid answers to the problems that exist in the company.

7.1.1 System and methodology

We are surrounded by systems. Our bodies are made up of various systems, such as the digestive system, the nervous system, etc. We form groups with other people in a social, political and economic way. At first glance, the various systems seem to have little in common. However, a closer look enables realizing that these phenomena are a collection of entities that are interconnected through defined relationships.

Elementary systems thinking emerged from the work of Gestalt psychologists, who emphasized the study of the mind as a whole rather than a collection of psychological units.

The idea of using the concept of system to understand phenomena is usually attributed to the work conducted in the 1930s by *Ludwig von Bertalany*, a German biologist. The general theory of systems is the name given to the discipline that formulates the principles applied to all systems. Following the pioneering works of *Bertalany*, systems thinking began to be applied to numerous fields and the Society for General Systems Research was created, including a group with *Bertalany, Rapaport, Boulding, and Gerard.*

It is consensual, nowadays, that the idea of system had a profound influence on the computer science field. The term "system" is part of the scientific basis of the area and is already part of the traditional view of the discipline. The disciplines of systems analysis and engineering contributed to develop the modern technological system, influencing the creation of the equipment that today we call computers. These disciplines also influenced the communications revolution we now call the Internet or "communication highways." The ability to integrate the organization into the system idea prevailed within the computer world.

The analysis of systems by *"hard"* methodology is concerned with determining the best way to achieve a certain objective, while the *"soft"* methodology system analysis intends to determine the objectives that are to be achieved.

Information systems use information technology and exist in a context of human activities. Work in the computer area is compatible with the *"hard"* and *"soft"* methodology. With regard to information technology, it is necessary to solve or improve the performance of the system and information technology (IT), which is a problem of *"hard"* systems. On the other hand, one must also solve or improve everything that concerns the effective use of the technological system and the issues related to human activities, which change with the application of a certain technology, this being a problem of the *"soft"* systems.

7.1.2 Methodologies in information systems

The term methodology does not have a clear definition, either in specific literature or among the professionals of the organizations. Moreover, there is very little consensus on its general meaning. The term is used in a very extensive way and with many interpretations, making it dicult to reach a consensus. Poor use of the term does not mean that there are no definitions; there are simply no universally recognized definitions. The methodology is generally defined as a series of steps and recommended procedures to be followed in the course of developing an information system.

The issues that appear are numerous and critical to the area of information systems. The scientific community of information systems regularly discusses the term methodology in the context of information systems, and so far it has not agreed on a universal definition.

The British Computer Society (BCS) developed one of the most commonly used definitions in 1983: "An information system methodology is a collection of philosophies, phases, procedures, rules, techniques, tools, documentation and training for information system workers (...) " referenced in Maddison (1983). Using this definition suggests that the methodology has a number of elements that should answer such questions as:

- How should the project be phased in?
- What tasks should be carried out at each stage?
- What are the outputs produced? When and under what circumstances should they be produced?
- What constraints should be applied?
- How should the project be managed and controlled?
- What support tools should be used?

In addition, the methodology should specify the training needs for its users and explain their philosophical foundations. This sometimes identifies unwritten aspects and reasons

that make the methodology an effective approach to the development of information systems from the author's point of view. The definition of a methodology should include a specific reference to its philosophy, in some cases, fundamental to understand a certain methodology. A methodology for information systems is much more than a series of techniques aided by software tools.

In practice, many methodologies, particularly commercial ones, are compact products that may include:

- Manuals
- Training and coaching (including videos)
- Consulting
- Pro forma documents
- Templates built-in in Templates, etc ...

Flynn and Diaz (1996) argues that the term methodology is not applied only in the context of information systems and that the term method is perfectly suitable to cover all that we mean by methodology. However, the same author states that "the term methodology was popular in the 1980s, but now it is no longer." The term methodology contains certain characteristics that the term method does not have, for example, the inclusion of "philosophy". Methodology is a broader concept than method.

7.1.3 The application model

The methodology starts by modelling the systems, applying some concepts discussed in Chapter 3 on simulation and system modelling. In this phase, it is fundamental to know the equipment well, as well as the processes that the company applies for its maintenance, the information system used, etc.

The application model is defined as an information collection system based on global models. It should be noted that companies often assume information collection systems, without first having defined and validated their maintenance model for a group of equipment, production line, etc. Then, the data is processed and analysed. This data is the source of the data for the simulation models to be tested. Scenarios are defined and then tested and compared. After that, it is decided if the model is correct or not. If is need to change the model, need to go back to the beginning. After the model is validated, the next step is to calculate the degree of importance of each component of the equipment, through several indicators. If this degree of importance has already been taken into account in maintenance management, then the methodology ends. Otherwise, it will be necessary to redefine the model and go back to the initial stage.

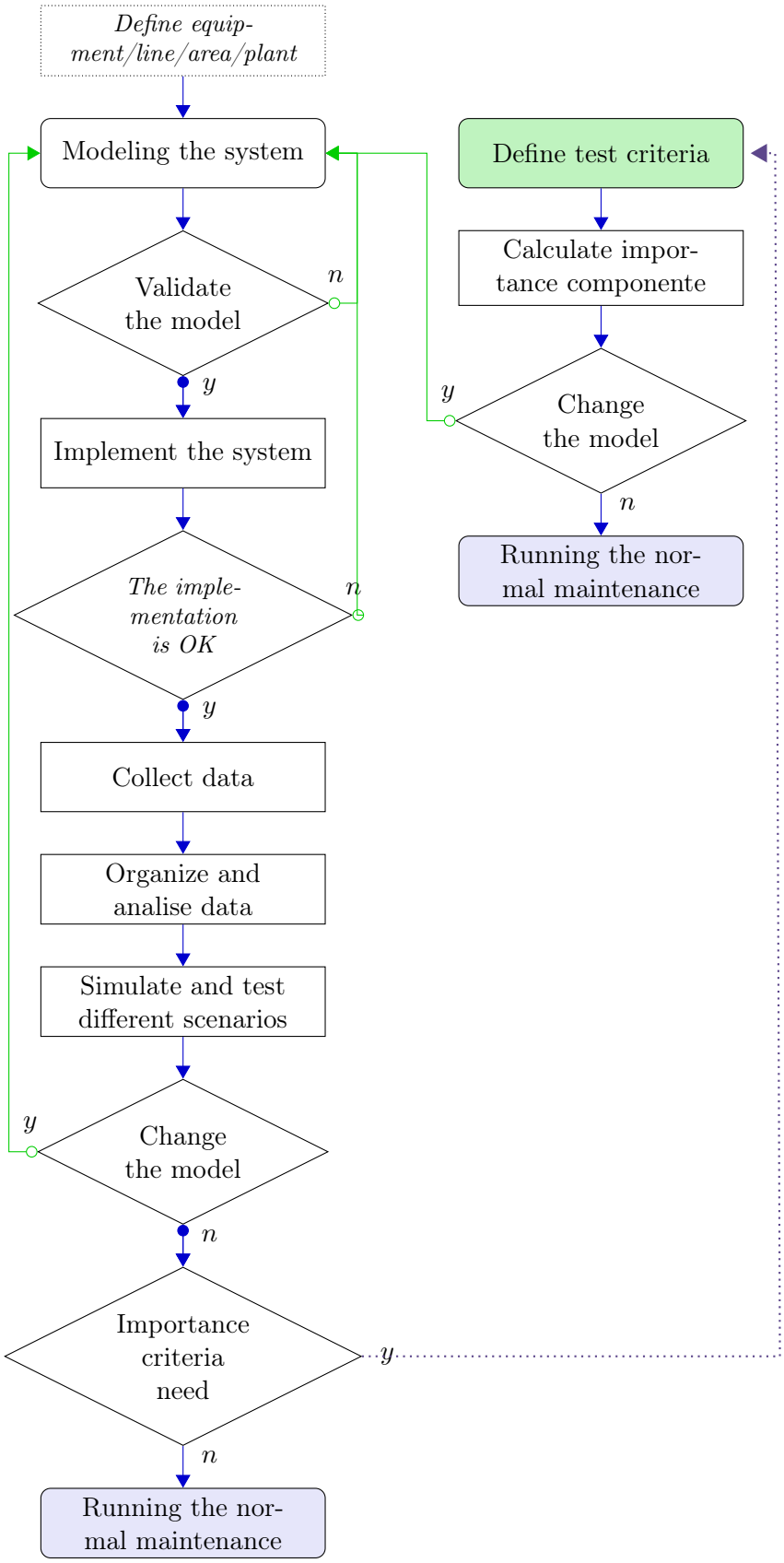


Figure 7.1: Methodological framework to apply component importance

7.2 Model the system

The work was carried out on a group of centrifugal pumps from a refinery that began operating in 1969. It is a hydro-skimming refinery with a distillation capacity of approximately 110 *kbpd*. The industrial complex also have an aromatic plant, a base oils plant, and a lubricants plant.

The annual output of the refinery aromatics plant is currently *440 kton*. Employing advanced technological systems, it excels in terms of reliability, economy of operation, and quality. The treatment capacity of the base oils plant is *1.18 mtpa* of crude oil. The lubricants plant has a production capacity of 80 kton per year. Figure 7.2 shows the refining process and its output.

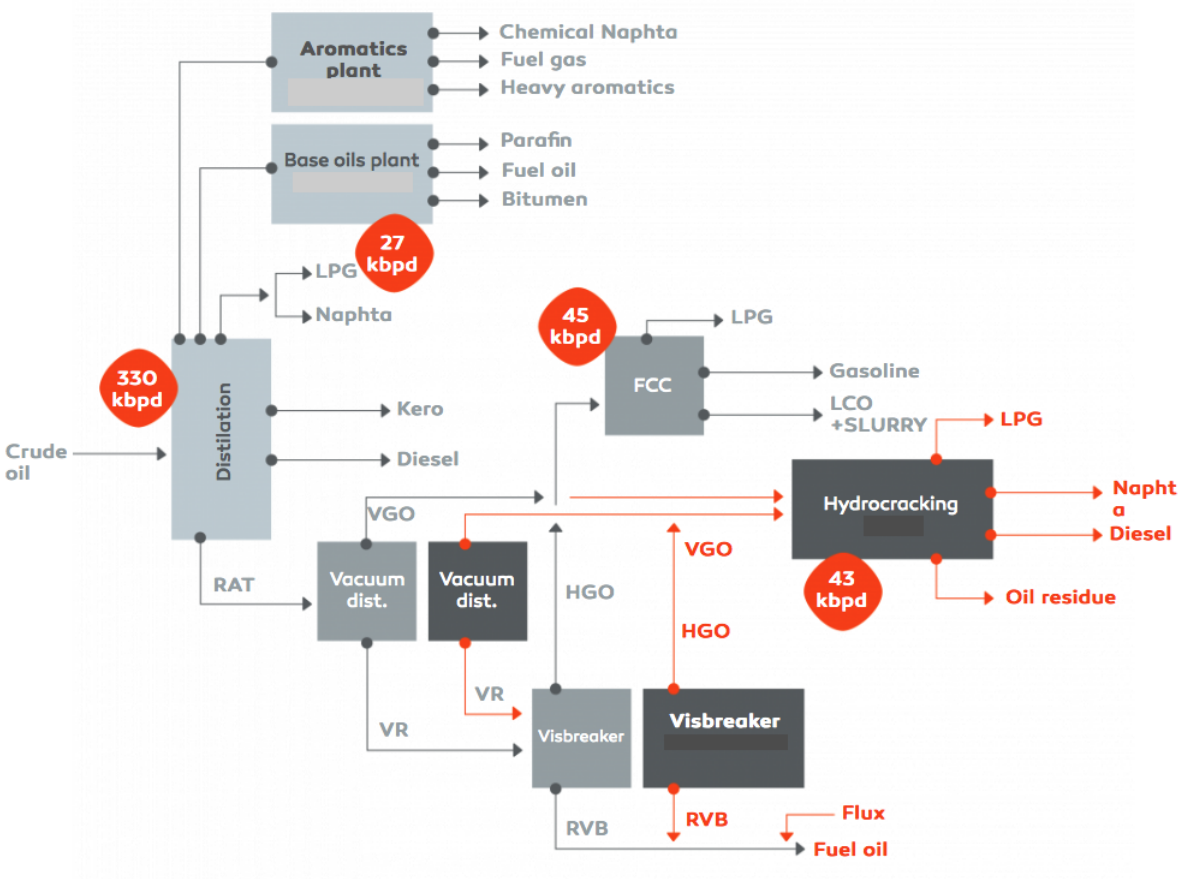


Figure 7.2: Oil refining system (Galp Corp. 2019)

The centrifugal pump is one of the most used electro-mechanical machines in the petrochemical industry. This type of pump is most used in the transport of fluids. It is estimated that among all pumps installed in the petrochemical industry, between 80% and 90%, are centrifugal pumps, Bloch and Budris (2004). Due to its excellent characteristics, the centrifugal pump is used for many applications, from water pumping to the transport of flammable fluids at high pressures and temperatures, Bloch and Budris (2004).

The centrifugal pump is very important equipment in the industrial process. It can be installed with a reserve pump in order to guarantee operational continuity in case of a failure of the main pump. Industrial hydraulic pumps have several characteristics that allow them to be differentiated and classified under different aspects. The most common classification divides hydraulic pumps into two large groups: dynamic pumps or turbo-pumps and positive displacement or volumetric pumps, Bloch and Budris (2004), Palgrave (2003). Centrifugal pumps are part of the first group. In turbo pumps, also referred to as dynamic pumps, acceleration is transmitted to the fluid so that it acquires kinetic energy from the transformation of the mechanical energy by means of the movement of the rotor inserted in the pump body.

The standard *ISO 14224:2016* provides a comprehensive basis for the collection of Reliability and Maintenance (RM) data in a standard format for equipment in all facilities and operations within the petroleum, natural gas and petrochemical industries during the operational life cycle of equipment. It describes data collection principles and associated terms and definitions that constitute a "reliability language" that can be useful for communicating operational experience. The failure modes defined in the normative part of this International Standard can be used as a "reliability thesaurus" for various quantitative as well as qualitative applications.

The standard *ISO 14224:2016* also defines the frontier of a pump within the petroleum, natural gas and petrochemical industries.

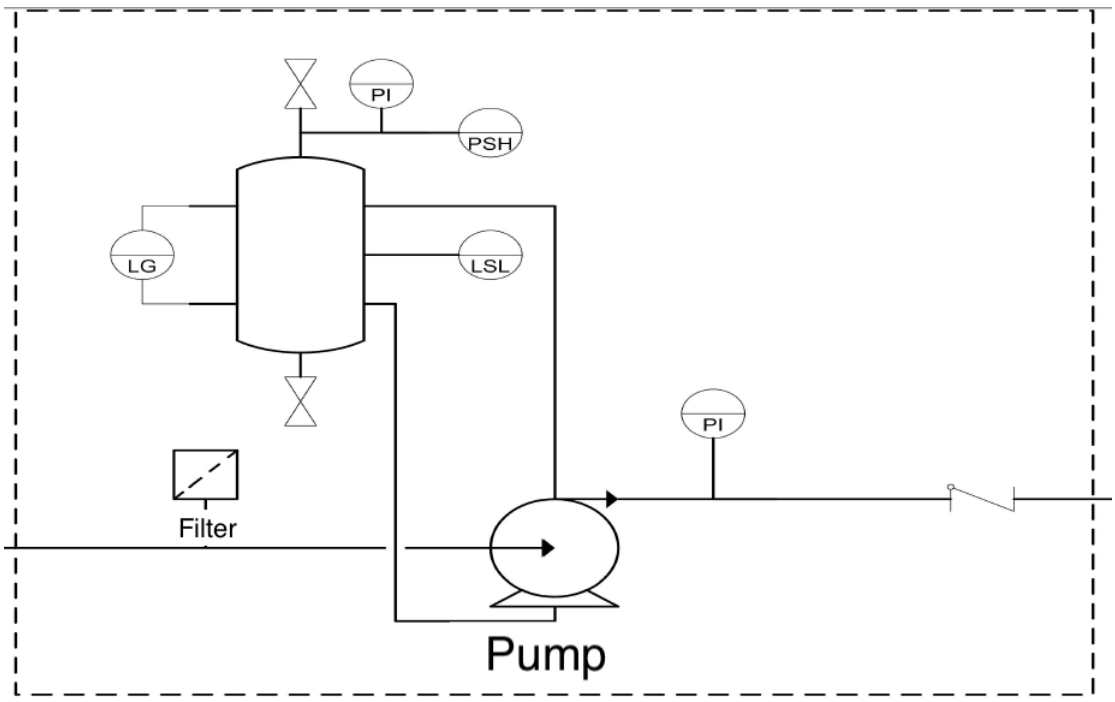


Figure 7.3: Frontiers of a pump (ISO 14224:2016)

The centrifugal pump is widely used because it has a simple operating principle, which

is materialized with constructive aspects that are easy to execute, has a high efficiency and is easy to operate and, if well applied, allows obtaining stable and good quality systems. Pumps are machines that transfer energy to the fluid for the purpose of transporting it according to process conditions. They receive energy from an external source, in this particular study, an electric motor, and give that energy to the fluid in the form of pressure, kinetic energy, or both, i.e., increase the pressure and/or velocity of the fluid. The movement of the fluid occurs by means of forces acting through the rotation of an axis coupled to the wheel (impeller, impeller) provided with blades in which it receives the fluid and conducts it through the periphery under the action of the centrifugal force.

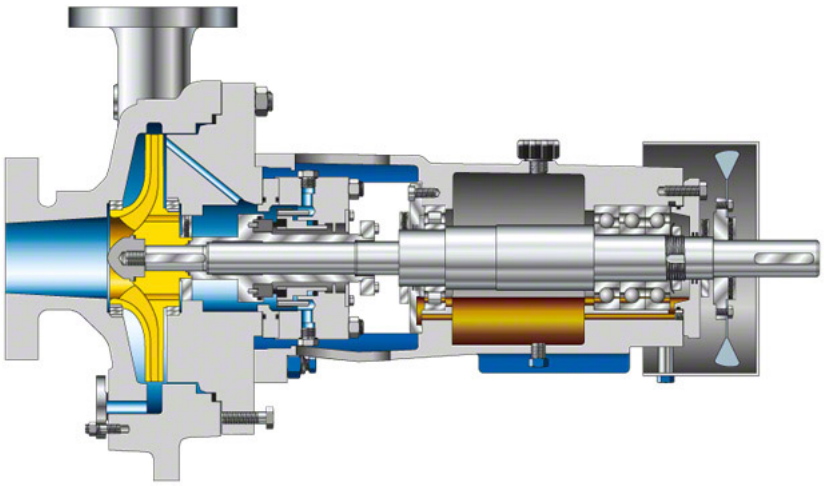


Figure 7.4: Centrifugal pump scheme (KSB corp. 2019)

Refinery pumps transport petroleum and its derived products in refineries, petrochemical plants and the chemical industry. They are used in temperature ranges from $-120^{o}C$ to $+450^{o}C$ at pressures of about *65 bar*. As the fluids handled are often highly volatile and flammable, the pump components in contact with the fluid handled are always made of ductile materials, such as unalloyed steel, chrome steel and, less frequently, nodular cast iron. The required Net Positive Suction Head - *NPSHR* value is particularly important and governs the selection of the drive speed and type of pump. Refinery pumps are most commonly single-stage horizontal volute casing pumps in back pull-out design.

There are many ways of describing, classifying and systematizing a centrifugal pump in the literature; this depends on the type of pump, its application, type of industry, liquid, etc. The classification adapted was Silva (2016), which defines the pump in the following scheme:

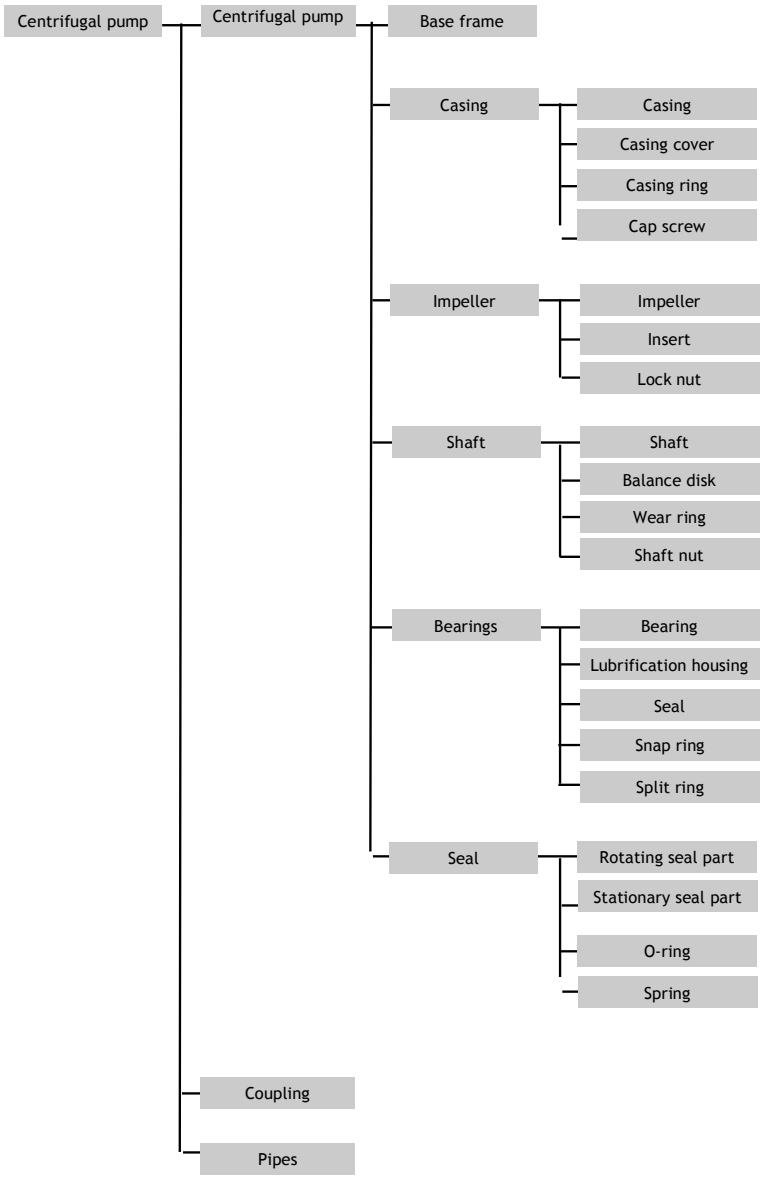


Figure 7.5: Schematic function - centrifugal pump (Silva 2016)

Analysing the different ways of representing and defining the block diagram of a centrifugal pump, and taking into account the type of maintenance that the company carries out and the grouping of faults/maintenance interventions, it has been defined for a five-block serial system with the following type of components:

Component	Description
C1	Main housing
C2	Impeller
C3	Shaft
C4	Bearings
C5	Seals

Table 7.1: Component descriptions of pump

A diagram block that represents a simplified model of the equipment under study is:

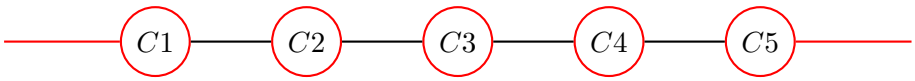


Figure 7.6: RBD structure diagram of a centrifugal pump

7.3 Data Failure analysis

The survey included data reported for seven centrifugal pumps, collected from 2001 to 2011. The failure data were collected only on centrifugal pumps in critical process. The reliability of these pumps is of great importance to the industry. The survey has limited the age of the pumps to no more than 10 years. The pumps move oil and diesel with different densities.

Many researchers use *OREDA (Offshore Reliability Data Handbook)* as a source of failure rate data to perform safety verification calculations. It remains an excellent reference for all who carry out data analysis.

Although the centrifugal pump populations in the two surveys are from different applications, it is possible to find many aspects of the *OREDA* report which helps our study. There are differences in the two studies, which makes a direct comparison of the results very difficult.

The pumps system is composed by two centrifugal pumps disposed in parallel. One pump is operating and the other pump is waiting to be demanded. The system for reliability analysis is called "Stand-by redundancy": just one pump is in operating mode and the other pump is waiting for failure of the first pump.

7.3.1 Inspection procedures for centrifugal pumps

Planned periodic inspections are performed to verify the system safety and performance by detecting potential and hidden failures and taking appropriate actions. If any problem is

Pump no.	ID Number	Type of Liquid
P1	1201	heavy liquid
P2	2015	heavy liquid
P3	2301	heavy liquid
P4	3701	aromatic
P5	3009	aromatic
P6	3011	aromatic
P7	3301	aromatic

Table 7.2: Type of liquid from pump

found during inspection, corrective actions are taken to restore the equipment to an acceptable level. In addition, a set of failure prevention actions may be taken to prevent future failures and restore equipment function. These include part replacement, re-tightening, lubrication, etc

Visual and operational checks of the centrifugal pumps safety and functionality are typically performed *6* times a day. This routine includes an inspection checklist for this particular class of equipment. The checklist contains, for example, actions to check a pump leaking and dripping through the seal, and unusual levels or types of noise. Vibration tests are performed once a month. Therefore, the time between the occurrence of the real failure and its detection is very short. Ignoring this time delay, we assume that the time between failures are complete data.

7.3.2 Proposals for the analysis data

The failure data analysis is carried out for different reasons: to identify weak areas in components, actions in maintenance and provide input for maintenance decisions. There is an important distinction to be made between components, that can only fail once and equipment and systems, that are repairable. Consequently, the analysis of failure data from populations of components or equipment can involve significantly different techniques.

The time interval between two consecutive failures is treated as complete data since assume that it is know the exact failure times. When the results of the last failures don't match with the end of the test, this is consider as right censored data.

As a first step it is always worthwhile calculating some parameters to summarize failure data. We measure the location of the observations and their dispersion around the mean value. For component failure data the measure of location is the mean life - Mean Time Between Failures (MTBF). The analysis of component failure data is too often undertaken before considering the pattern of failure exhibited by the dataset. Summary statistics provide some indication of the properties of individual data samples but on their own have this limitations.

The preliminary analysis data discussed here can provide a better understanding of the likely failure patterns, the symmetry, range and modality of a dataset. Any one of these

analyses will be an improvement on summary statistics and will probably indicate where further investigation is likely to be worthwhile. In data collection, it is possible to see some asymmetry. All data have some considerable standard deviation. This is because the range of data for all the pumps is quite large and the number of events for each pump is few.

In the seven pumps, one had very little data and was not representative of the normal "failure" pattern of the other pumps, so it was decided to remove the data from the study and at a later stage analyse the reason why its behaviour was totally different from other pumps. For our study, their inclusion would greatly bias the analysis and there is probably a reason for this anomalous behaviour.

Therefore, because of its flexibility, the exponential model will be used to analyse the typical industrial failure data. It is particularly useful for small data when there is uncertainty regarding the fit to other distributions.

7.3.3 Results - data analysis

The first result that was obtained was the *MTBF* by equipment and, then, the overall *MTBF*, not forgetting that there are two groups of pumps with different maintenance due to the density of the fluid which they operate. The maintenance costs are significantly different from one group of pumps to another (heavy and light), as can be seen in table 7.3. In this analysis, and clearly in a real analysis, there are many factors that influence reliability, productivity and costs. From table 7.4 and table 7.3 it is possible to conclude that the relation of the cost of maintenance with the *MTBF* is not linear. The heavy pumps have a lower average cost and are those which have lower *MTBF*. This is because the cost is more diluted by a greater number of failures than is included in the group of light pumps.

Pumps no.	MTBF (u)	Mean Cost ($)
1201	1638	2829,17
2015	842,94	1383,98
2301	1884,23	1604,90
3009	1456	96,27
3011	2016	878,029
3301	3573,81	18,75

Table 7.3: MTBF and pump cost

Pump Group	Mean Cost ($)
B1	1939,35
B2	331,02

Table 7.4: Mean cost of the two groups of pumps

With the survey performed, it is possible to have a reasonable number of data that makes the mean values obtained meaningful. Thus, the failure rates per component, according to the following block diagram representation, are:

	Pump n0.							
Comp	**1201**	**2015**	**2301**	**3009**	**3011**	**3301**	MTBF (u)	Hazard Rate (1/u)
c1	15724,8	16016	96096	14560	39312	78624	43388,8	0,000023
c2	13104	6864	7392	0	26208	26208	13296	0,000075
c3	78624	6406,4	24024	0	78624	39312	37831,73	0,000026
c4	6048	1656,82	6406,4	2569,41	4624,941	11232	5422,93	0,000184
c5	3418,4	4576	5338,66	4368	4914	8736	5225,1	0,000191

Table 7.5: MTBF and hazard rate for each components

7.4 Simulation - results

The simulation developed three scenarios: the first scenario is the simple case of a series system designed by *5-out-of-5*. In the second and third scenarios, the faults are provided in consecutive blocks and, therefore, consider the case in the system have two consecutive components fails and the system fail, that is the scenario *2-out-of-5:F* or, have only two components working to the system continues to functions, the *2-out-of-5:G* scenario.

Sample	Quarter	Semi-Annual	Annual
10	0.4	0.2	0
100	0.35	0.07	0.01
500	0.352	0.114	0.012
1000	0.346	0.12	0.015
2000	0.362	0.112	0.011

Table 7.6: Simulation reliability of 5-out-of-5

The results of the simulation show that the reliability for a time interval that has a range from *3* months to one year decreases in all cases. And when the sample number increases, the values stabilize around a value that will be the closest to the reliability of that model.

Clearly can be concluded that the reliability of the consecutive F and G series systems are much higher than the simple series system. In practice, what happens is that there is usually not a single repair, but a set of two or three repairs (to simplify the system, the study approach two consecutive systems). The results are in the table 7.8 and are quite elucidative.

After this analysis, a simulation was developed to calculate costs based on the reliability of each system. In this case, considering an acceptable reliability, given that the system is

Sample	Quarter	Semi-Annual	Annual
10	0.8	0.6	0.4
100	0.85	0.68	0.22
500	0.86	0.616	0.232
1000	0.869	0.607	0.248
2000	0.864	0.6245	0.293

Table 7.7: Simulation reliability of a consecutive 2-out-of-5:F system

Sample	Quarter	Semi-Annual	Anual
10	1	0.8	0.6
100	0.97	0.92	0.56
500	0.944	0.846	0.566
1000	0.954	0.842	0.575
2000	0.941	0.838	0.572

Table 7.8: Simulation reliability of a consecutive 2-out-of-5:G system

redundant, of no lower than 90%, our calculation provided that for scenario *1* of the simple serial system it would be necessary to carry out two maintenance actions per quarter and for, the other two scenarios, a single intervention is sufficient. If one compares the annual cost of each option, it will necessarily have different maintenance policies and different budgets.

	Annual Cost	
	B1	B2
5-out-5	11636.13	1986.12
2-out-5:F	5818.06	993.06
2-out-5:G	5818.06	993.06

Table 7.9: Simulation Costs Scenarios

Reliability analysis is easier when done early on in the project and the tests are performed in the laboratory under controlled conditions. When equipment goes to the field and it is necessary to take into account diverse variables that influence reliability, it is no longer as easy to use or choose the model and, many times, the data itself is not as reliable. Actually, modelling in a laboratory or project is much simpler and easier to do. That's why our methodology aims to help those who want to analyse and improve the reliability of their equipment that is in operation.

7.4.1 Results - component importance

The case study also developed the component importance calculation, for different time horizons: quarterly, semi-annual and annual. Four criteria have been chosen, which have provided very interesting results. The action and use of this information depends on the organization and its planning and definition of the maintenance policy and strategy. The components that are most important need, of course, more preventive maintenance, inspection routines, etc.

	Import.Birbaum					**Improvement Potential**					**Risk Reduction Worth**					**Criticality Import.**				
Time	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
T	0.001	0	0.001	0	0	0.001	0	0.001	0	0	1.001	1	1.001	1	1	0.001	0	0.001	0	0
S	0.009	0.003	0.008	0.002	0.002	0.009	0.003	0.008	0.002	0.002	1.009	1.003	1.008	1.002	1.002	0.008	0.003	0.007	0.001	0.001
A	0.062	0.023	0.055	0.014	0.014	0.061	0.023	0.055	0.014	0.014	1.066	1.024	1.059	1.014	1.014	0.051	0.019	0.046	0.012	0.011

Table 7.10: Component importance Simulations - 5-out-of-5

In this case, for the three proposed scenarios, there are differences in the criteria, which can also help us to better understand which components are really important or not. In all the scenarios, components *C4* and *C5* appear the most important, although their importance is differences from scenario to scenario and time horizon.

	Import.Birbaum					**Improvement Potential**					**Risk Reduction Worth**					**Criticality Import.**				
Time	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
T	0.024	0.338	0.055	0.148	0.007	0.023	0.321	0.052	0.14	0.006	1.025	1.511	1.058	1.174	1.007	0.024	0.34	0.055	0.149	0.007
S	0.044	0.561	0.109	0.26	0.014	0.036	0.469	0.091	0.218	0.011	1.045	2.278	1.122	1.352	1.014	0.047	0.607	0.118	0.282	0.015
A	-0.022	0.748	0.059	0.375	-0.004	-0.014	0.455	0.036	0.228	-0.002	0.978	3.964	1.062	1.6	0.996	-0.03	1.007	0.079	0.505	-0.006

Table 7.11: Component importance Simulation - 2-out-of-5:F

	Import.Birbaum					**Improvement Potential**					**Risk Reduction Worth**					**Criticality Import.**				
Time	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
T	0.126	0.077	0.311	0.364	0.308	0.109	0.067	0.269	0.315	0.266	1.144	1.084	1.452	1.571	1.445	0.138	0.085	0.342	0.4	0.338
S	0.171	0.106	0.347	0.565	0.477	0.109	0.068	0.221	0.36	0.304	1.206	1.119	1.532	2.299	1.913	0.243	0.151	0.494	0.803	0.678
A	0.136	0.092	0.219	0.692	0.581	0.038	0.026	0.062	0.195	0.164	1.157	1.102	1.281	3.25	2.388	0.395	0.269	0.638	2.013	1.69

Table 7.12: Component importance Simulation - 2-out-of-5:G

7.4.2 Conclusion

The final goal of the case study under analysis is not a classical approach to reliability, with estimation and adjustment of data, but to validate a simulation methodology that can contribute to a reliability analysis and that can be a tool for the definition of maintenance policies, defining resources, and promoting maintenance management with the use of simulation techniques.

For a complete, accurate and thorough analysis of reliability, another type of data and collection procedure is required. In fact, data is used for reliability, but they serve to illustrate the models and methodologies, as simulation proposals. The data was collected for sufficient number of years and from a reasonable number of almost identical centrifugal pumps but, in the future, to carry out a better reliability analysis, more accurate data will be necessary.

The work identifies the advantages of simulation in the field of reliability, and demonstrates that its application is possible and useful.

For this, it is necessary to master a set of techniques and knowledge in the technical area (in this case, pumps), as well as predictive maintenance techniques (vibration analysis, etc.), statistics (data analysis and estimation of parameters), modelling of systems and processes, and simulation (GRN, software, etc.).

For the case study, the data collect and analysed are good enough, because they fit into average values that, together, can have very reasonable and acceptable average values for the simulation.

The time to failure of the most critical pumps were gathered from computer maintenance systems. The accuracy of the results depends on the total number of failures and their completeness or censoring status. The results are better and more consistent with more failure data and, mainly, with complete failure data.

The analysis of the data from of an oil refinery pumps with great period of working time is not simple because of the amount of censored and missing information and of the relative equipment ageing process.

The results of the statistical analysis could be used to develop the optimum inspection interval for each pump. Different trends and values have been obtained for different pumps and a model should take into account the failure trend and life pattern. The analysis should be performed carefully, since the same equipment in the same refinery may exhibit different failure patterns depending on the operating and the environmental conditions, namely the viscosity of the fluid and the age. With this analysis, operations management can make the right decision in advanced to avoid any critical failures and plants downtime.

Chapter 8

Conclusions and future work

Today, in our companies we have connected all equipment, which becomes intelligent, even autonomous, which allows to be informed at any time and guided in our choices to give power of decision to the equipment. The evolution of technology, has enabled the development of new tools and methods to promote condition monitoring or preventive maintenance and avoid failures and accidents.

Simulation studies use computer intensive procedures to assess the performance of a variety of statistical methods in relation to a known truth. Designing high-quality simulations that reflect the real situations and complex equipment seen in practice, is not a simple process. All simulation studies involve the generation of several independent simulated data sets. These generated data sets must also be completely independent for the different scenarios considered, such as in the presence of censored data.

The thesis intend to contribute in the design and programming algorithms that generate correctly, robust and non-skewed censored data and are a useful tool in the field of simulation on reliability.

The dissertation begins, in chapter 2, by theoretically approaching the concepts and describing the basic concepts of reliability analysis in engineering and maintenance. Chapter 3 starts by providing a brief introduction to system simulation and modelling. A summary of the state of the art in the generation of random numbers, as well as the hypothesis test are then provided to identify whether the data obtained could be considered as random. In Chapter 4, the censored data type is defined and the algorithms to simulate the different censored data are provided, as is a simulation and comparison of the five statistical distributions. In Chapter 5, the estimation models for the various types of censored data are developed; the EM method is also developed to right type I censor data with Weibull distribution.

Chapter 6 begins with the revised theory about reliability of complex and coherent systems, *RBD* and *FTA*. Then, simulation studies are presented for the various cases. Several Monte Carlo simulation methods are explained and simulated. The last part of the chapter presents the reliability of linear consecutive *k-out-of-n*, component importance

and simulations studies.

Chapter 7 starts off with the methodology for the use of the simulation tools in the reliability of equipment, in order to manage the maintenance of a company (or its improvement and optimization), and proceeds with an investigation case carried out on a set of critical equipment.

Finally, this last chapter presents a summary of the work developed and of the main conclusions, as well as the future perspectives within the thematic area in which we develop the dissertation.

8.1 Conclusions and goal achieved

The objectives that were defined at the beginning of the dissertation were carried out through bibliographical research of the state of the art, by the adaptation of simulation tools to support maintenance management using the information systems of the companies.

In a transition from an academic environment to the working world, a subject was chosen: simulation on reliability, that only started to have more prominence in the last decade. As a result of this development, the simulation has also been supported by increasingly versatile and comprehensive information systems, fully embracing today's computer development. On the other hand, we also see, more recently, that maintenance has an increasingly prominent role in organizations, leaving the traditional view of generating costs.

Finally, and particularizing a type of industry, a refinery plant served as a case study. Maintenance contributes very actively in a sector recognized for the high quality and safety requirements of the product. These were the themes combined - simulation and maintenance - developed throughout the work, supported by their theoretical bases, and put into practice. The possibility of analysing real maintenance data was the perfect scenario for the thesis project, and the fulfilment of the expectations about the work.

Throughout the work and in the main chapters *(4 to 7)*, the results were analysed and the appropriate conclusions were reached. Therefore, a more global summary is present in line with the proposed objectives:

1. The thesis was used to develop reliability model simulation algorithms for complex equipment/systems and when data collection is faced with censored data;

2. The algorithms are innovative and their development was carried out using three different software programs: - *Python, Matlab and R*; which allowed verifying and comparing some characteristics, such as: performance, ease of programming and debugging, versatility and available functions, etc.;

3. The document proposes a methodology and approach for data analysis and application of simulation tools in industry companies;

4. The method of maximum likelihood and also the EM method were developed, in order to analyse censored data and have more precision and confidence in the results;
5. A methodology of analysis (hypothesis tests) and validation with an evaluation matrix is proposed to test *i.i.d.* data of the RNG of censored data;
6. A sequence of preventive interventions could be similar to *k-out-of-n* consecutive linear reliability models by analogy and mathematical modelling;
7. The methodology developed is a mixture of several academic models, tools and simulation methods, as well as of the organizational structure of the companies;
8. There are authors - few in our view - who mention the importance of aspects of simulation in maintenance; hence, in maintenance, there is a very diffuse notion of management and information systems;
9. Reliability analysis enables a reflection on the current system of company equipment and together with simulation tools contributes to show, in a very positive way, the actions of change that significantly improve an organization's maintenance service;
10. The work for the modelling of an equipment or system should involve a multidisciplinary team, that will accompany the development of all the work;
11. The results of the simulation are quite positive, but can be boosted with the introduction of new approaches in the industry, such as, machine learning, artificial intelligence, etc.;
12. The execution and application time of modelling and simulation is usually long (this is one of the disadvantages of this methodology) because it requires the involvement of a good number of an organization's employees, the availability of their time, data collection and validation of results;
13. Investment in training, in technical areas of maintenance, together with simulation work, is profitable and can bring many benefits to the productive sector and lead to a good organizational environment;
14. It is important, within academic work in the engineering, management and information systems areas, to establish or to take into account the definitions that are assumed regarding the concepts of modelling, simulation and reliability;
15. In Portugal, there is not much work on maintenance management, much less on maintenance simulation;
16. For analysis and simulation work on reliability or maintenance, as well as for the implementation of an information system with integrated simulation, it is necessary to define a methodology.

8.2 Future work

After working within the scope of the project, it was possible to identify some improvement areas, thus, making it possible to provide a proposal for future work:

1. Apply the methodology proposed in Chapter *7* to other equipment and in other organizations;
2. Extend simulation methodology and tools to the maintenance logistics sector;
3. Introduce new parameters and analysis with the use of machine learning tools and algorithms;
4. Improve and automate the data collection process by developing a routine or even a software module that provides automatic reporting. Although ambitious, this intensive and possibly high-investment programming software, shall undoubtedly provide an added value to the entire information management system of enterprise;
5. For further analysis, it would also be very interesting to link the data collected with production, quality and logistic data;
6. Comparative and benchmarking studies of the algorithms applied in the different software;
7. Adapt the simulation models for integration in *DES* (discrete event system) and *ABM* (agent based modelling) simulation models;
8. Integrate simulation models into the most current models of risk analysis;
9. Compare the methodology proposed with other methodologies and tools for the implementation of information systems in industrial maintenance;
10. Study the importance of simulation models in maintenance management;
11. Study the impact of integrating production and maintenance information systems in the automation systems for the future of maintenance management;
12. Further develop the modelling theory for simulation in maintenance.

Appendix A

Monte Carlo methods simulation on reliability

In reliability the use of Monte Carlo methods - MCm began at least with Orkand (1960). Can be divided in two major groups: the methods to repairable systems and the methods to non-repairable subsystems.

Two methods to non-repairable sub-systems and one method to repairable systems were select and develop in *matlab* the algorithms of each method to simulate. Not all of algorithms have the same results and precision that have the results made in the original works and this can be, because was used different and modern tools to generate random numbers and resolve numerically the calculus.

The first method is named *Kamat-Riley* method - *K-R*, from Kamat and Riley (1975) and assume that individually the subsystem is independent of each other, and life distribution is known for each subsystem and distribution parameters have been estimated. In the beginning of the method it's necessary to determine all minimal tie-sets. The *K-R* method using the normal approximation to the binomial distribution will result in some error. The s-confidence intervals obtained by *K-R* method can be shrink by increasing number of simulation cycles. he *K-R* method can be accepted in large number of engineering applications.

The second method is named *Rice-Moore* method - *R-M* from Rice and Moore (1983) that used the technique fail-pass failure. The *R-M* method can be applied to any complex system structure but the subsystem failures have to follow binomial distributions. The method is applied especially to the system with zero-failure subsystem. The *R-M* simulation method is also based on the normal approximation to the binomial distribution. The *LCLs* obtained by the *R-M* algorithm are larger than the exact *LCLs*.

The third method named Kim-Lee Monte Carlo method - *KLMC* from Kim and Lee (1992) use the relationship between the system MTBF and the component failure processes, to predict the exact value of *MTBF*. It's necessary to know all the sample paths of the component failure processes. The *KLMC* algorithm can be applied to any binary coherent system with known component lifetime distributions.

A.1 The method Kamat-Riley

This Monte Carlo procedure, developed by Kamat and Riley (1975), can be applied to most systems with arbitrary system reliability structure and subsystem with different failure distribution. In this method, individual subsystem are assumed to be independent of each other and repair of failed subsystem are not allowed; the underlying life distribution is known for each subsystem and statistical distribution parameters have been estimated.

According to Kamat and Riley (1975) the procedure has several steps that are:

(a) Find out all minimal tie-sets form system Reliability Block Diagrams (RBD). Assume that we need to obtain system reliability interval estimates at some point t.

(b) From the life distribution of each subsystem, a random failure time t_i, is generated where I represents the ith subsystem, 0<i<n.

(c) Compare t_i, with t for all subsystem. If $t_i > t$, this indicates that at time t subsystem I functions properly; if $t_i \leq t$, then subsystem I has failed.

(d) Determine whether the whole system is functioning or down according to the states of its subsystem at t from step (c). Check all subsystem in a minimal tie set. If all of them are operational, then the system operates properly as t. If one or more of them fail, then the tie–set is broken (failure) at t. Further, check next minimal tie-set until an unbroken one appears, which means that the system is a operational at t. If all minimal tie-sets are broken, then the system fails at t.

(e) Repeat steps (b), (c), (d) for, say, n times. Count failure and success numbers of the system respectively: *ns(t)* and nf(t). Note that $n = ns(t) + nf(t)$

(f) The system reliability point estimation corresponding to t is given by

$$R(t) = \frac{n_s(t)}{n_s(t) + n_f(t)} \tag{A.1}$$

Note that the simulation results are of binomial type. Based on the normal approximation to the binomial distribution, the $100(1-\alpha)\%$ confidence intervals of system reliability at time t are given by

$$[R_l(t), R_u(t)] = R(t) \pm Z_y \frac{R(t)(1 - R(t))}{[n_s(t) + n_f(t)]^{\frac{1}{2}}} \tag{A.2}$$

Where Z_y is the double side $100(1-\alpha)\%$ percentile of the standard Normal distribution with mean zero and variance 1.

Example A.1.1 (Example method Kamat-Riley)**:**
An application example is given by Kamat and Riley (1975). The system reliability structure diagram in this example is show in figure (A.1) and lifetimes of all nine subsystem: *a,b,c,d,e,f,g,h,i* are assumed to follow the two-parameter Weibull distribution with survival function

$$s_f(t; K_i, M_i) = e^{\left[-\frac{K_i}{M_i+1} t^{M_i+1}\right]} \;\; and \;\; t, K_i > 0, \; M_i > -1 \tag{A.3}$$

From figure (A.1), can see that it is difficult to determine the reliability interval estimates of this system by using classical statistical. Per system reliability theory, the system's minimal tie-sets can be found to be: *adg,bdg,adhi,bdhi,aefi,befi,cfi.*

The scale parameter K_i and shape parameter M_i values for all nine subsystems are listed in (A.1). Using the *K-R* Monte Carlo algorithm, **1000** simulation replications are performed; the table (A.2) show the results of system reliability point estimates at certain time points.

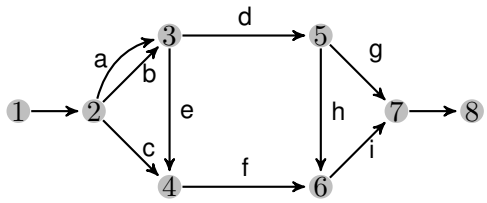


Figure A.1: Reliability structure diagram

Component no.	Scale parameter K	Shape parameter M
A	2.8	1.8
B	2.7	1.7
B	2.6	1.6
D	2.5	1.5
E	2.4	1.4
F	2.2	1.2
G	2.3	1.3
H	2.1	1.1
I	2	1

Table A.1: Weibull parameters for each component

Only a part of the program made in Matlab is show, and to illustrate the most important part of the main routine. The program obtain directly the results that are in the table (A.2).

```
            % step that compare if the system have a failure or not
            % create the vector that if the sum is equal 53 represents that
            % the equipement not failure
            tabela = t*trec;
            sumato = sum(tabela);
                if sumato == 53;
                    ns = ns +1;
                else
                    nf = nf +1;
                end
        end
    rt(j)= ns/(ns+nf);
    %confidence interval
    Rl(j)=rt(j)-1.645*((rt(j)*(1-rt(j)))/sqrt(ns+nf))
    Ru(j)=rt(j)+1.645*((rt(j)*(1-rt(j)))/sqrt(ns+nf))
end
```

Algorithm 13: Program Kamat and Riley (partial)

Kamat and Riley (1975) do not derive exact 95% confidence intervals for this system by classic statistical methods and not discuss the accuracy of their simulation result. The K-R method using the normal approximation to the binomial distribution will result in some error. From (A.2) can see that the confidence intervals are generally very narrow and

the point estimates are at the middle of them.

Time	Lower 2.5% limit	Reliability point estimate	Upper 2.5% limit
0	1	1	1
0.1	0,988964	0,990099	0,991234
0.2	0,93889	0,943894	0,948899
0.3	0,880513	0,888614	0,896715
0.4	0,800701	0,811881	0,823061
0.5	0,714482	0,727723	0,740963
0.6	0,646665	0,660537	0,67441
0.7	0,578853	0,592822	0,606791
0.8	0,517766	0,531353	0,54494
0.9	0,471231	0,484158	0,497086
1	0,427983	0,440144	0,452305

Table A.2: Reliability simulation results - 95% confidence intervals

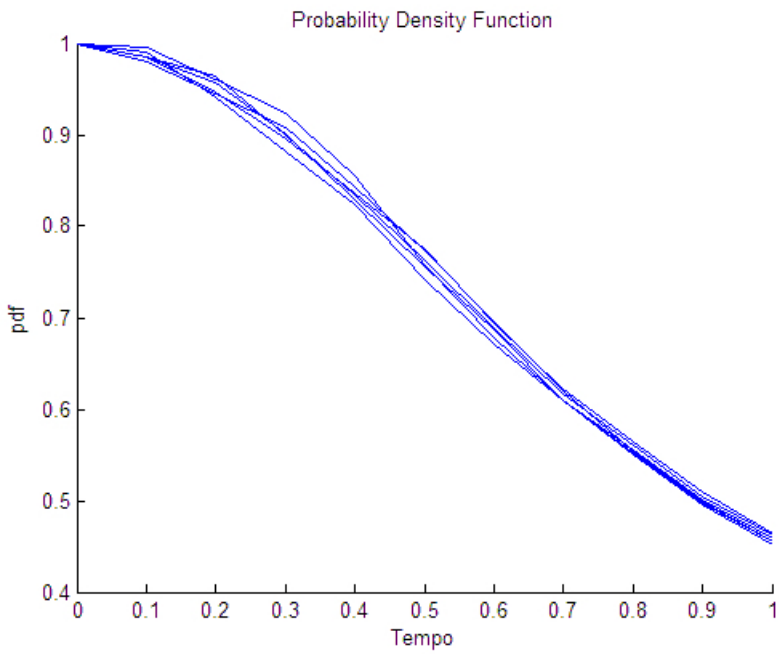


Figure A.2: Example from Kamat and Riley

A.1.1 Development and explanation of Kamat Riley algorithm

Start by defining the matrix (*tie* in the algorithm) that defines the critical path: - if it is only in series equipment (let's assume 4 components) the matrix stays:

$$M = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$$

The *ncc* (number of critical paths is 1) which is equal to m representing the number of rows in the array, and *ccc* (maximum number of components in all cc) that is equal to the number of columns in the array.

If it is two parallel lines with *3* components each the array looks like this:

$$M = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

In the second step of the cycle is generated the vector with the times of each component: vector T_i

$$Ti = \begin{array}{c} \begin{array}{cc} Componente & tempo \end{array} \\ \begin{pmatrix} 1 & 0.9 \\ 2 & 1.5 \\ 3 & .5 \\ 4 & 3 \end{pmatrix} \end{array}$$

From this two matrices will be generated another matrix F by comparison with the simulation time - t that will be incremented each time depending on the *for* statement:

```
for(t in seq(0, 0.1, by = 0.01) )
```

The matrix F is constructed with the comparison $t_i < t$ where if true F_i, j takes the value of zero and if False takes the value 1. Below the construction of the matrix and an example for the case of equipment with two parallel branches with three components each.

$$M = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \qquad Ti = \begin{array}{c} \begin{array}{cc} Component & time \end{array} \\ \begin{pmatrix} 1 & 0.9 \\ 2 & 1.5 \\ 3 & .5 \\ 4 & 3 \\ 5 & 7 \\ 6 & 2 \end{pmatrix} \end{array} \qquad Ci = \begin{bmatrix} 0.9 & 1.5 & .5 \\ 3 & 7 & 2 \end{bmatrix}$$

Another way of doing the algorithm starts by constructing another matrix in which it replaces the number of the component in the matrix *tie* by the value of the failure time. From there it would be to compare for each element of the matrix whether the value is $<$ or $>$ that the t (simulation time).

$$F_{i,j} = \begin{cases} 0 & se\, T_i\{M_{i,j}\} < t \,(\text{failure component}), \\ 1 & se\, T_i\{M_{i,j}\} \geq t \,(\text{component working}). \end{cases} \qquad F_{i,j} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

From the matrix F we can construct the vector g that informs us for each critical path if it is stopped or to function, through the following function:

$$G_i = \begin{cases} 0 & se \sum_{i=1}^{m} F_m = 0 \,\text{CC it's working,} \\ 1 & se \sum_{i=1}^{m} F_m \geq 1 \,\text{CC it's stopped.} \end{cases} \quad G_i = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Finally, the vector r is constructed through the matrix G. If the $\sum G = m$ is equal to the total number of critical paths, that is the value m, it means that all critical paths are stopped and therefore the equipment has failed or takes value 0 if $\sum G \neq m$ some critical path is working and soon the machine is working.

$$R_i = \begin{cases} 0 & se \sum_{i=1}^{m} G_m = m \;\; \text{Equipament stopped} \\ 1 & se \sum_{i=1}^{m} G_m \neq m \,\text{Equipament working.} \end{cases} \quad G_i = \begin{bmatrix} 0 \\ 1 \\ 1 \\ \cdots \\ \cdots \\ 1 \end{bmatrix}$$

A.2 The method Rice and Moore

The Rice and Moore (R-M) simulation method is also based on the normal approximation to the binomial distribution and have the follow parameters: the success probability of a binomial or the reliability of component is p, failure probability q, number of components n, failure number f_i, then this binomial failure follows the normal distribution with mean p and a variance *(pq/n)*.

If a component has zero failure, that is, $f_i = 0$, then f_i can be replaced by the equivalent failure number f_i given by *Gatliffe (1976)*.

The Key steps of the R-M process are:

1. Define the system and its Reliability Block Diagram (RBD). Develop the algorithm to compute system reliability from its subsystem reliabilities, i.e., system reliability structure function.

2. For each subsystem, determine its failure fi or equivalent failure number fi'. To simplify, they are both represented by fi *.

3. Calculate estimates:

$$p_i = 1 - \frac{f_i^*}{n_i} \;\; q_i = 1 - p_i$$

$$Asymptotic\ variance = \frac{p_i q_i}{n_i}$$

4. For each subsystem, generate a random variable form *N(0,1)*, where *N(0,1)* is the normal distribution with mean zero and variance 1.

5. Find the second estimate $p \sim (p_i, \frac{p_i q_i}{n_i})$ by drawing an r.v. from *N(0,1)*. Multiply by asymptotic standard variance and add it to p_i

6. Calculate system reliability R, from subsystem reliability according to the algorithm created in step (a).

7. Implement steps (d)-(f) many times for, say, *999* times.

8. List these values in order of no n-decreasing magnitude.

9. Determine the 100(1- α)% percentile to obtain 100(1- α)% LCLs of Rs

Here we will show the programming code made in Matlab. From this program we directly obtain the results that is show in the figure (A.3).

Figure A.3: Example from Rice and Moore with f1=1 and f2=2

A.3 The method KLMC

The MTBF is often used as a measure of repairable-system reliability.

Prediction of MTBF is an important aspect of the phase design of systems. The MTBF is used as a measure of repairable-system reliability. The MTBF of a system can be predict before the equipment is made but its necessary to know the relationships between the system MTBF and the component failure processes. The MTBF is defined by ET, where T is the time to failure of the equipment, but this definition assume independent and identically distributed random failures.

Kim studied the relationship between the system MTBF and the component failure processes. The KLMC algorithm is applied to any binary coherent system with known component lifetime distributions.

As stated by Kim and Lee (1992), to *KLMC* method. it's necessary to do some *assumptions*:

1. The system has n components. Each component is either operating or failed, and so is the system.

2. Each component lifetime distribution is known.

```
%variis
n= 10 %numero de amostras
k = 2 % numero de componentes
mu = 0; % the mean
sigma = 1; % the standard deviation
%rt = 1
f(1) = 1
f(2) = 3

for v=1:10
for j = 1:1:1000

 for i = 1:k;

     p(i) = 1- f(i)/n
     q(i) = 1-p(i)
     AV (i) = p(i)*q(i)/n

     z =  unifrnd( mu , sigma);
     pf (i) = p(i) + z*(AV(i).^(0.5));


 end

 rt (j)  = pf (1)* pf (2)

end

 X= sort(rt,'ascend');

 rc(v) = X(50)
 end

 hold on

 plot (rc)
```

Algorithm 14: Program Rice and Moore

3. The state of the system is determined solely by the states of the components by means of a binary s-coherent structure function.

4. The states of the components are s-independent.

5. Failed components are replaced (with new ones) at system failure.

6. If an MPS (minimal path set) fails, components therein (which are not included in any other MPS) cease to operate until repair of the system.

7. Replacement time for any component is negligible. Replacement/repair does no damage the system.

Assumption *6* is more realistic that to separate maintenance where an operating component can still fail even though it is in a failed MPS.

Notation

i	index for components; i = 1,2, ..., n
j	index for MPSs; j = 1,2, ..., mi
k	index for inter-failure times; k = 1,2, ...
m_i	number of MPSs that contain component i
$P_{i,j}$	MPS j which contains component i
q_k	inter-failure time (of system) k; $q^0 = 0$
t_i^k	residual lifetime of component i at system failure k - 1
$L_{i,j}^k$	inter-failure time k of $P_{i,j}$
S_i^k	$max_j\{L_{i,j}^k\}$

Remarks

1. Since $P_{i,j}$ is a series system, then $L_{i,j}^k = mint_i^k$ over all $s \in P_{i,j}$ for fixed k
2. Since the system fails when the last MPS fails, $q^k = max_{i,j}\{L_{i,j}^k\}$ over all j for each i, over all i for fixed k.
3. If $t_i^k \leq S_i^k$, then component i has failed during $[q^{k-1} - q^k]$.
4. If $t_i^k \geq S_i^k$ then component i is used for S_i^k during the interval $q^{k-1} - q^k]$ and has residual life $t_i^{k+1} = t_i^k - S_i^k$

KLMC Algorithm

1. Find $P_{i,j}$ for all i j.
2. Initialize: k=l. Generate t_i^k for all i; k is fixed (MTBF = "Undefined").
3. $L_{i,j}^k = min_s\{t_s^k\}$ over all $s \in P_{i,j}$ for fixed k.
4. $q^k = max_{i,j}\{L_{i,j}^k\}$ over all j for each i, over all i for fixed k.
5. $S_i^k = max_j\{L_{i,j}^k\}$ over all j for each i, for fixed k.
6. If $t_i^k \leq S_i^k$ then generate t_i^{k+1} else $t_i^{k+1} = t_i^k - S_i^k$.
7. If $\sum_{i=1}^k q^i \leq t$ then set k = k + l , and go to step 3.
8. If $k > 1$, then MTBF = t/(k-1).

For a given time t, the algorithm produces the *MTBF* estimate by dividing t by the number of failures in *[0,t]* under the condition that separated maintenance is not considered.

A part of the programming code made in Matlab is show. This program obtain directly the results that is show in the figure (A.5).

```
% i - index for components
% j - index for MPS
% k - index for inter failure times
% mi - numero of mps que contem a componente i
% pij - MPS j que contem a componente i
% qk - inter failure time do sistema k; qk(0)= 0
%tik - residual life of compoenente i at system failure k-1
%lij - inter-failure time K of pij
% Sik - max(lij)

clear all

%Step 1 - Find pij for all i&j
%Definir a matriz que contem a minimal path sets
pij = [1 1 0;1 0 0;0 1 0 ;0 1 0 ; 0 0 1 ];
dist= [1 0.1 0;2 10 2;3 2 2;1 0.1 0;4 2 2]; % fun distribui
[n,ff]=size(dist);
 re=10 % replicas
 f=1
 t=200

for t=10:10:200

    for s=1:1:re

h=0
q=0
k=1 % initialize K

%Step 2 - Generate life times
   for i=1:n
       tgi =  gertempo (dist, i)
       tg(i)=tgi
   end
 %Matriz MPS por colunas
     for i=1:n
            for j =1:ff
                pji(i,j) = pij(i,j)*tg(i)
```

Algorithm 15: Program KLMC (partial)

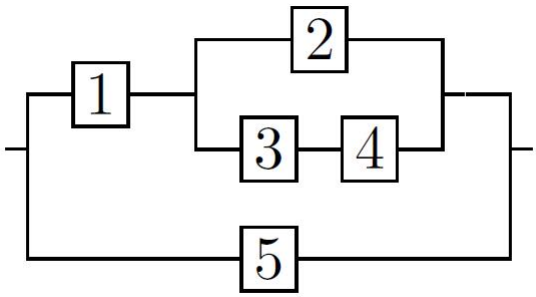


Figure A.4: A 5 s-independent component coherent system

Example A.3.1 (The method KLMC)**:**
A system has 5 s-independent components with minimal path sets $\{1,2\},\{1,3,4\},\{5\}$ as in *A*.4. The lifetime *CDF's* for each component are given. Calculate *MTBF* (5O).

The components have the following lifetime CDF's:

1. expf (0. 1t) - exponential
2. gaufl (t-10) /2] - Gaussian
3. weif(2t; 2) - Weibull
4. expf(0. 1t) - exponential
5. gamf(2t; 2) - Gamma

Algorithm Steps

1_1. $P_{1,1} = \{1, 2\}$, $P_{1,2} = \{1, 3, 4\}$
$P_{2,1} = \{1, 2\}$
$P_{3,1} = \{1, 3, 4\}$
$P_{4,1} = \{1, 3, 4\}$
$P_5 = \{5\}$

2. Generate the lifetimes for each components
$t_1^1 = 10$, $t_2^1 = 15$, $t_3^1 = 35$, $t_4^1 = 45$, $t_{5]^1=15}$

3_1. $L_{1,1}^1 = L_{2,1}^1 = min\{10, 15\} = 10$
$L_{1,2}^1 = L_{3,1}^1 = L_{4,1}^1 = min\{10, 35, 45\} = 10$
$L_{5,1}^1 = 15.$

4_1. $q^0 = max\{10, 10, 15\} = 15$

5_1. $S_1^1 = max\{10, 10\} = 10$, $S_2^1 = 10.$
$S_3^1 = 10$, $S_4^1 = 10$, $S_5^1 = 15$

6_1. Since $t_1^1 = 10 = S_1^1$ then component 1 has failed;
$t_1^2 = 37$ is the generated lifetime.
Since $t_2^1 = 15 > 10 = S_2^1 = t_2^2 = 15 - 10 = 5$
Since $t_3^1 = 35 > 10 = S_3^1 = t_3^2 = 35 - 10 = 25$
Since $t_4^1 = 45 > 10 = S_2^1 = t_4^2 = 45 - 10 = 35$
Since $t_5^1 = 15 = S_5^1$ then component 5 has failed;
$t_5^2 = 20$ is the generated lifetime.

7_1. $q_1 = 15 < t = 50$; $set\, k = 1 + 1 = 2$; go to step 3

3_2. $L_{1,1}^2 = min\{37, 5\} = 5 = L_{2,1}^2$
$L_{1,2}^2 = min\{37, 25, 35\} = 25 = L_{3,1}^2 = L_{4,1}^2$
$L_{5,1}^2 = 20.$

4_2. $q_2 = max\{5, 25, 20\} = 25$

5_2. $S_1^2 = max\{5, 25\} = 25$.
$S_2^2 = 5$, $S_3^2 = 25$, $S_4^2 = 25$, $S_52 = 20$.

6_2. Since $t_1^2 = 37 > 25 = S_1^2$, then $t_1^3 = 37 - 25 = 12$.
Since $t_2^2 = 5 = S_2^2$, then component 2 has failed;
$t_2^3 = 18$ is the generated lifetime
Since $t_3^2 = 25 = S_3^2$ then component 3 has failed;
$t_3^3 = 22$ is the generated lifetime.
Since $t_4^2 = 35 > 25 = S_4^2 \, then \, = t_4^3 = 35 - 25 = 10$
Since $t_5^2 = 20 = S_5^2$ then component 5 has failed;
$t_5^3 = 11$ is the generated lifetime.

7_2. $q^1 + q^2 = 15 + 25 < t = 50; setk = 2 + 1 = 3$; go to step 3

3_3. $L_{1,1}^3 = min\{12, 18\} = 12 = L_{2,1}^3$
$L_{1,2}^3 = min\{12, 22, 10\} = 10 = L_{3,1}^3 = L_{4,1}^3$
$L_{5,1}^3 = 11$.

4_3. $q^3 = max\{12, 18\} = 12$.

5_3 and 6_3. [not needed].

7_3. $q^1 + q^2 + q^3 = 15 + 25 + 12 > t = 50$.

8_3. MTBF(50) = 50/(3-1) = 25.

Blank periods appear at $10 \leq t \leq 15$ for components 2, 3, 4, since (by assumption 6) these components go into "suspended animation" after the failure of component 1 at t = 10. Figure A.5 is the result of the simulated MTBF (5000 simulations) for t = 10 to 200 for this example 5-component system.

The validity of this simulation can be checked by comparing the simulation results with the theoretical ones. Figure (A.5) shows that the simulated *MTBF* converges to a limit. This limit should agree with the analytic *MTBF* for t goes to infinity, or at least lie above the lower bound for $r > 0$, if the *KLMC* algorithm is correct.

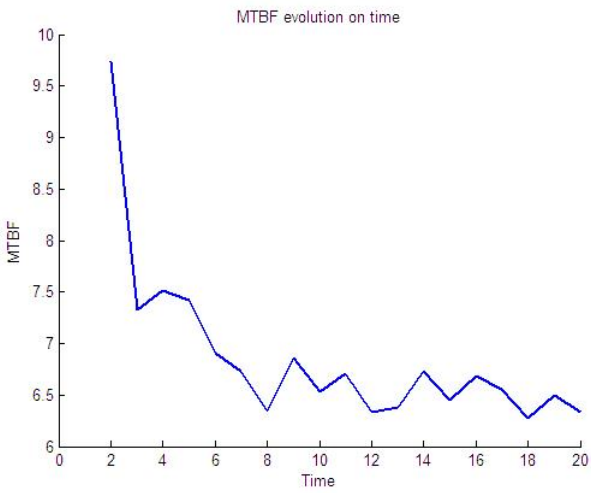


Figure A.5: MTBF for complex system

A.4 The Monte Carlo method to availability of systems

The availability of systems with independent components, proposed from Dubi (2000) can be obtained by normalized system function. It is desired to have a general solution by which, for any non-normalized $S(\overrightarrow{B})$, a general solution can be obtained. For large systems, the derivation of normalized system function can become very difficult, and a non-normalized system function is relatively easy to obtained.

A general algorithm is obtained based on a non-normalized system function, with the following consideration: the sate space contains 2^n different state vectors of the form $\overrightarrow{B}_j = (b_1^j, b_2^j, \ldots, b_n^j)$ *with* $j = 1, \ldots, 2^n$; b_i can take the values 1 and 0 and $p_i = P_r[b_i = 1]$. Denoting

$$\beta_i = \begin{cases} \text{if } b_i{=}1 \\ \text{if } b_i{=}0 \end{cases}$$

the probability that the state vector is $\overrightarrow{B} = (b_1, b_2, \ldots, b_n)$ at time t (since p_i may be time dependent), is given by

$$P_r(\overrightarrow{B}, t) = \prod_{i=1}^{n} \beta_i \tag{A.4}$$

Let ϕ_s denote the set of state vectors $\{\overrightarrow{B}\}$ for which the system is operational, i.e. $\phi_s = \{\overrightarrow{B} : S(\overrightarrow{B}) > 0\}$. The complementary of ϕ_s,ϕ_f, is defined as $\phi_f = \{\overrightarrow{B} : S(\overrightarrow{B}) = 0\}$. Since the state vectors are mutually exclusive, the probability of the event ϕ_f is the sum of the probabilities of the state vectors included in it. Thus, the unavailability can be express

in the form:

$$U(t) = 1 - A(t) = P_r\left[\overrightarrow{B},(t)=0\right] = P_r[\Gamma_f] = \sum_{\substack{j=1 \\ \overrightarrow{B}_i \in \Gamma_f}}^{2^n} P_r(\overrightarrow{B},t) = \sum_{\substack{j=1 \\ S(\overrightarrow{B}_i)=0}}^{2^n} \left(\prod_{i=1}^{n} \beta_i^j(t)\right) \tag{A.5}$$

Equation (A.5) lead us to a very simple algorithm for the calculation of the unavailability, namely:

1. Scan through all 2^n state vectors.

1.1 For each state vector $\overrightarrow{B}_j$,check if $S(\overrightarrow{B}_j) = 0$.

1.2 If condition 1.1 is fulfilled, calculate the state probability according to expression (A.4) and add it to the unavailability, otherwise proceed to the next state vector.

This procedure is realized in the program *Matlab*. The algorithm requires a scan over all 2^n vector states. Table A.3 shows the results of calculations for a serial system with n components, where n varies from 15 to 21. The availability of each component is $p_i = p =$ 0.998, i=1,...,n. Indeed, this is a trivial case. This could be calculated as $U = 1 - p^n$, but the algorithm is intended for any system function. The number of different state vectors, and therefore the calculation time, depended only upon n (the dimension of the problem), and is independent of the particular system function used.

Number of components	**Unavailability**	**Number of vector states**
15	0.029583	32768
16	0.031523	65536
17	0.03346	131072
18	0.03539	262144
19	0.03732	524288
20	0.03924	1048576
21	0.04116	2097152

Table A.3: Calculation of the unavailability

A.4.1 Solution by Monte Carlo

As stated by Dubi (2000) the phase space is defined by the vectors $\overrightarrow{B}_j = (b_1^j, b_2^j, \ldots, b_n^j)$. Define the estimator as the random variable:

$$\eta(\overrightarrow{B}_j) = \begin{cases} 1, & S(\overrightarrow{B}_j) = 0 \\ 0, & \text{otherwise.} \end{cases}$$

The mean value of $\eta(\overrightarrow{B}_j)$ is given b

$$E(\eta) = \sum_{j=1}^{2^n} P_r(\overrightarrow{B}_j, t)\eta(\overrightarrow{B}_j) = \sum_{\substack{j=1 \\ S(\overrightarrow{B}_j)=0}}^{2^n} P_r(\overrightarrow{B}, t) = U(t). \tag{A.6}$$

Thus $\eta(\overrightarrow{B}_j)$ is an unbiased estimator of U(t). To sample a state vector, discrete sampling is used. For i=1,...,n, a random number ξ_i is compared to pi. If $\xi_i \leq p_i(t)$, then$b_i = 0$. The estimation of U(t) using N histories takes the form:

1.Use N histories, J=1,...,N.

1.1 In each history sample a state vector $\overrightarrow{B}_j$.

1.2 If $S(\overrightarrow{B}_j) = 0)$ set $\eta(\overrightarrow{B}_j) = 1$, otherwise set $\eta(\overrightarrow{B}_j) = 0$.

2.Upon completion of N histories, the estimate of the unavailability is

$$U = \frac{1}{N}\sum_{j=1}^{N} \eta(\overrightarrow{B}_j) \tag{A.7}$$

with the PRSD given by

$$\xi = 100 \times \sqrt{\frac{1-U}{N \times U}} \tag{A.8}$$

In table *A.4* results for the same problem as that describes in table *A.3* are displayed for system with up to 300 components. In this particular problem the unavailability increases with the number of components, therefore the Percentage Relative Standard Deviation - PRSD, decrease.

A.4.2 Variance reduction for the availability

The reason for the high variance systems, is because the event of interest is rarely sampled in this case a state vector for which the system is failed. Variance reduction methods are based on seeking a different set of probabilities for the occurrence of the events, the process is also referred to as "biasing".

Consider expression (*A.6*) for the unavailability

$$E(\eta) = \sum_{j=1}^{2^n} P_r(\overrightarrow{B}_j, t)\eta(\overrightarrow{B}_j)$$

Number of components	Unavailability estimate U	Exact result U	PRSD (%)
5	0.00983	0.00996	3.13
15	0.0291	0.0295	1.82
30	0.0581	0.0582	1.27
40	0.0774	0.0769	1.1
50	0.0946	0.0952	0.98
70	0.1297	0.1307	0.82
90	0.1655	0.1648	0.71
100	0.1821	0.1814	0.67
300	0.3293	0.3229	0.45

Table A.4: Results of Monte Carlo calculations on a serial system

Let K denote the number of zeros in the vector $\overrightarrow{B}$, i.e. K is the number of failed components. Let Λ_k be the subset of all state vectors with exactly k failed components. The number of state vectors in Λ_k is $\binom{k}{n}$. The averaging of expression (A.6) can be written as

$$E(\eta) = \sum_{j=1}^{2^n} \eta(\overrightarrow{B}_j) P_r(\overrightarrow{B}_j, t) = \sum_{k=0}^{n} \sum_{\overrightarrow{B} \in \Lambda_k} \eta(\overrightarrow{B}_j) P_r(\overrightarrow{B}_j, t) \tag{A.9}$$

The second moment, $S^2(\eta) = E(\eta)$, because $\eta(\overrightarrow{B})$ is a binomial estimator. Biasing is applied by changing the state probabilities using a new set of probabilities, $P_r^*(\overrightarrow{B}_j, t)$, instead of $P_r(\overrightarrow{B}_j, t)$. To keep the estimator (the first moment) unchanged (unbiased), the expression is written as

$$E(\eta) = \sum_{k=0}^{n} \sum_{\overrightarrow{B} \in \Lambda_k} \left(\eta(\overrightarrow{B}_j) \frac{P_r(\overrightarrow{B}_j, t)}{P_r^*(\overrightarrow{B}_j, t)} \right) P_r^*(\overrightarrow{B}_j, t) = \sum_{k=0}^{n} \sum_{\overrightarrow{B} \in \Lambda_k} (\eta(\overrightarrow{B}_j) W(\overrightarrow{B}_j)) P_r^*(\overrightarrow{B}_j, t) =$$

$$= \sum_{k=0}^{n} \sum_{\overrightarrow{B} \in \Lambda_k} (\eta^*(\overrightarrow{B}_j) P_r^*(\overrightarrow{B}_j, t)) \tag{A.10}$$

This can be compared to the multiplication and division by a new PDF done for integrals estimations.

$$W(\overrightarrow{B}_j, t) = \frac{P_r(\overrightarrow{B}_j, t)}{P_r^*(\overrightarrow{B}_j, t)} \tag{A.11}$$

is called a weight. In the above process, the analogue probabilities are replaced by the biased probabilities. To maintain the estimator unbiased the analogue estimator $\eta(\overrightarrow{B}_j)$ is multiplied by a weight $W(\overrightarrow{B}_j)$ yielding a new estimator $\eta^*(\overrightarrow{B}_j)$. While maintaining the first moment unchanged, it is desired to change the second moment so that it is smaller then analogue second moment, leading to a reduction of the variance. The second moment

takes the form

$$S^{*2}(\eta) = \sum_{k=0}^{n} \sum_{\overrightarrow{B} \in \Lambda_k} \left(\{\eta(\overrightarrow{B}_j) \frac{P_r(\overrightarrow{B}_j, t)}{P_r^*(\overrightarrow{B}_j, t)} \right)^2 P_r^*(\overrightarrow{B}_j, t) = \sum_{k=0}^{n} \sum_{\overrightarrow{B} \in \Lambda_k} (\eta(\overrightarrow{B}_j) \left(\frac{P_r(\overrightarrow{B}_j, t)}{P_r^*(\overrightarrow{B}_j, t)} \right) P_r(\overrightarrow{B}_j, t) \tag{A.12}$$

Optimum values for the biased probability set can be obtained using Lagrange multipliers similar to the method used for integration with piecewise constant distributions. Taking the derivative of the second moment with the constraint of the normalization of the biased probabilities yields

$$\frac{S^{*2}(\eta)}{dP_r^*(\overrightarrow{B}_j, t)} = \frac{d}{dP_r^*(\overrightarrow{B}_j, t)} \left\{ \sum_{j=1}^{2} {}^{n}\eta(\overrightarrow{B}_j) \left(\frac{P_r^2(\overrightarrow{B}_j, t)}{P_r^*(\overrightarrow{B}_j, t)} \right) + \lambda \left(\sum_{j=1}^{2^n} P_r^*(\overrightarrow{B}_j, t) - 1 \right) \right\} = 0 \tag{A.13}$$

yielding

$$-\eta(\overrightarrow{B}_j) \left(\frac{P_r^2(\overrightarrow{B}_j, t)}{P_r^*(\overrightarrow{B}_j, t)} \right) + \lambda = 0 \ or \ P_r^*(\overrightarrow{B}_j, t) = \sqrt{\frac{\eta(\overrightarrow{B}_j) P_r(\overrightarrow{B}_j, t)}{\lambda}} \tag{A.14}$$

Upon substitution to the normalization condition, and using the fact that $\eta(\overrightarrow{B}_j)$ is a binomial estimator, one obtain the optimal probabilities in the form

$$P_r^*(\overrightarrow{B}_k)_{opt} = \frac{\eta(\overrightarrow{B}_k) P_r(\overrightarrow{B}_k, t}{\sum_{j=1}^{2^n} \eta(\overrightarrow{B}_j) P_r(\overrightarrow{B}_j, t)} \tag{A.15}$$

Optimal biasing is obtained when the probabilities are proportional to the true importance for each point in the space.

Substituting the optimal probabilities back in (*A*.12) have the same effect as in the case of optimal biasing of integrals.

To apply biasing and improve the efficiency *(FOM)* of the calculation, some approximation assumptions must be used.

A.4.3 Single parameter biasing

Assuming that the components have identical availabilities, namely $p_i(t) = p \ \forall \ i = 1, ..., n$ the probability of any state $\overrightarrow{B} \in \Lambda_k$ is $p_k = (1-p)^k p^{n-k}$. The probability depends only upon the number of failed components in the state. Expression (*A*.11) then takes the form

$$E(\eta) = \sum_{k=0}^{n} (1-p)^k p^{n-k} \left[\sum_{\overrightarrow{B} \in \Lambda_k} \eta(\overrightarrow{B}_j) \right] = \sum_{k=0}^{n} (1-p)^k p^{n-k} A_k \tag{A.16}$$

where

$$A_k = \sum_{\overrightarrow{B} \in \Lambda_k} \eta(\overrightarrow{B}_j)$$

is the number of states in Λ_k which results in a failed system. Let $p^* =$ with $\alpha \leq 1$. Reducing the availability of each component will increase the probability of subsets Λ_k of higher K to appear, thereby increasing the probability of sampling a failed state vector with the above formulation $P_r^*(\overrightarrow{B}) = (1-p)^k p^{*(n-k)}$.

Substituting in (A.10) and (A.12) yields:

$$E(\eta) = \sum_{k=0}^{n} (1-p)^k p^{n-k} \left[\sum_{\overrightarrow{B} \in \Lambda_k} \frac{(1-p)^k p^{(n-k)}}{(1-p^*)^k p^{*(n-k)}} \eta(\overrightarrow{B}_j \right] \tag{A.17}$$

and

$$S^{2*}(\eta) = \sum_{k=0}^{n} \sum_{\overrightarrow{B} \in \Lambda_k} \eta(\overrightarrow{B}_j \left(\frac{(1-p)^k p^{(n-k)}}{(1-p^*)^k p^{*(n-k)}} \right) (1-p)^k p^{(n-k)} =$$

$$= \sum_{k=0}^{n} \frac{(1-p)^{2k} p^{(n-k)}}{(1-\alpha p)^k \alpha^{(n-k)}} \Lambda_k \tag{A.18}$$

To realize this biasing technique, sampling of components states is done with a probability $p^* = \alpha p$. The estimator $\eta(\overrightarrow{B}_j)$ is multiplied, for each component, by (1-p)/(1-p*) if the state is failed, or by p/p* if the state is not failed. The accumulated weight will become

$$W(\overrightarrow{B}_j) = \frac{(1-p)^k p^{(n-k)}}{(1-p^*)^k p^{*(n-k)}} \tag{A.19}$$

in accordance with equation (A.17). By estimating Λ_k the second moment can be calculated as function of α, and an optimal value of α can be estimated. For the serial case discussed above, $A_0 = 0$ and for k > 0 $A_k = \binom{n}{k}$. Since all vector states with at least one failed component are failed states, the second moment takes the form

$$S^{*2} = \sum_{k=1}^{n} \binom{n}{k} \frac{(1-p)^{2k} p^{(n-k)}}{(1-\alpha p)^k \alpha^{(n-k)}} = \left(\frac{\alpha - 2\alpha p + p}{(1-\alpha p)\alpha} \right)^n - \left(\frac{p}{\alpha} \right)^n \tag{A.20}$$

P	**Unavail U**	α **Optimal value**	$S_{*2}(\eta)$ **biased** 2nd **moment**	σ_0 **unbiased** $\sqrt{U(1-U)}$	σ_0^* $\sqrt{S^{*2}(\eta) - U^2}$	$B_e =$ $\sigma_0^2 \tau / \sigma_0^{*2} \tau$
0.9	0.40951	0.87	0.2884	0.4917	0.3475	1.99
0.99	0.04901	0.81	0.005637	0.2158	0.05687	13.7
0.999	0.00499	0.8	6.055E-05	0.0704	0.00597	139
0.9999	0.0004999	0.8	6.098E-07	0.0223	0.00059	1387
0.99999	4,999E-0.5	0.8	6.10E-09	0.0070	6,002E-05	13683

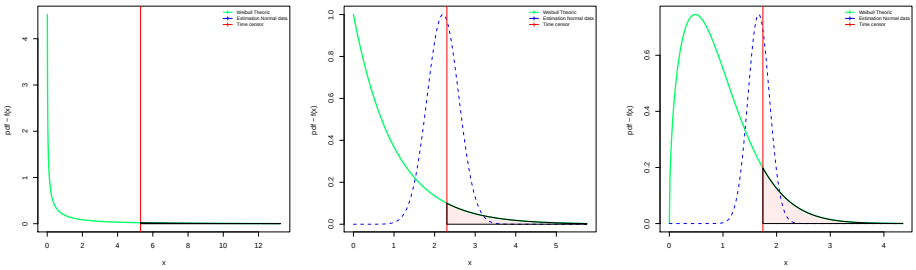
Table A.5: Expected benefit at optimal single parameter biasing

The optimal results for the five components serial case with p=0.9, 0.99, 0.999, 0.9999 and 0.99999 are summarized in table *A*.5. The benefit ratio β_e is defined as the ratio of the FOMs (Figure of Merits) of the biased and unbiased (analog) cases. It indicates the time ratio of the two methods in achieving any given *PRSD*.

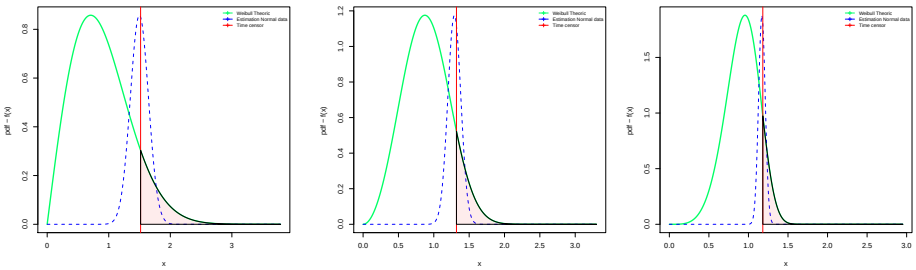
Appendix B

Simulation data censored with statistical distribution

B.1 Figures from Weibull distribution

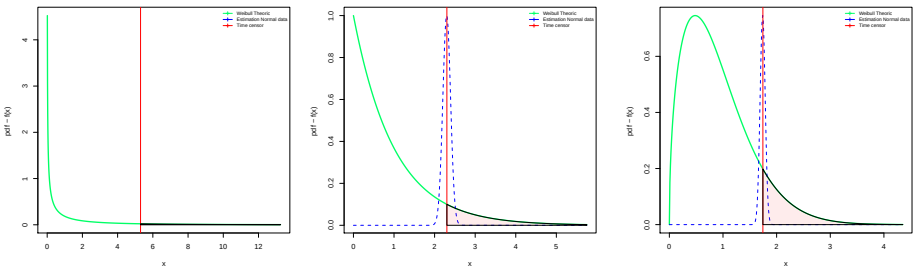


$\beta_{0.5}$ β_1 $\beta_{1.5}$

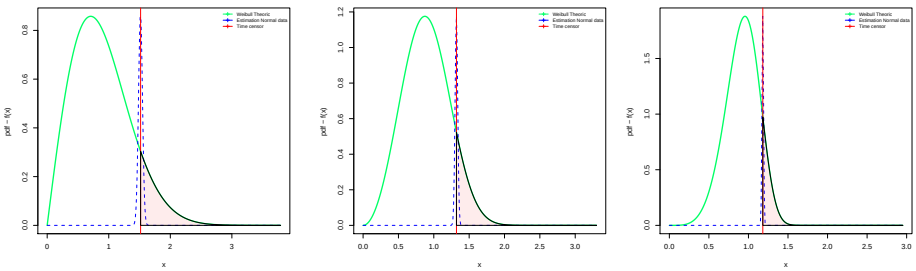


β_2 β_3 β_5

Figure B.1: Simulation right type I, Weibull (β), $C = 10\%$, $n = 50$



$\beta_{0.5}$ β_1 $\beta_{1.5}$



β_2 β_3 β_5

Figure B.2: Simulation right type I, Weibull(β), $C = 10\%$, $n = 1000$

B.2 Figures from Gamma distribution

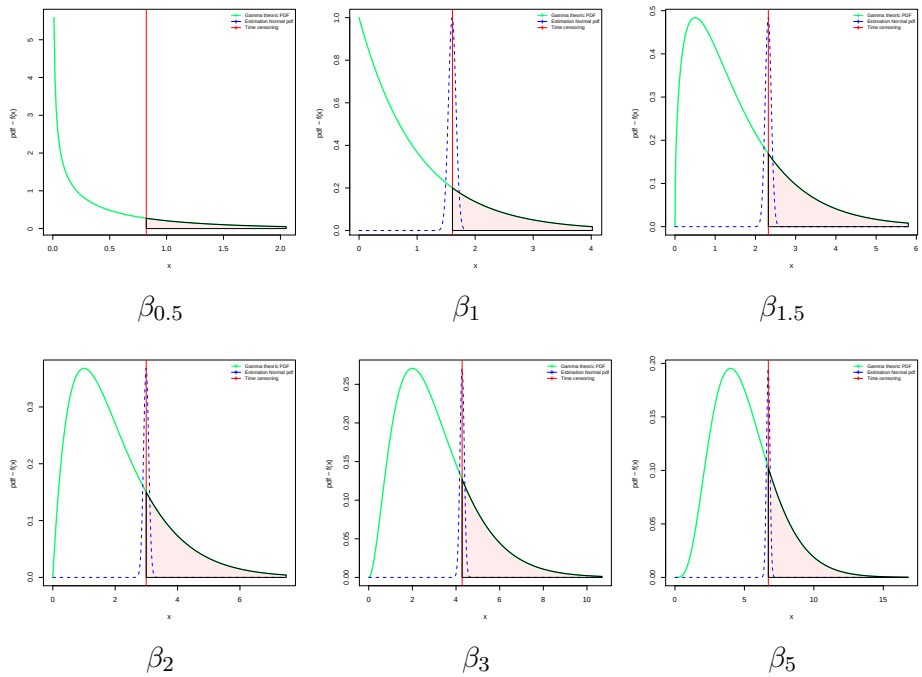


Figure B.3: Simulation right type I, Gamma (α), $C = 20\%$, $n = 50$

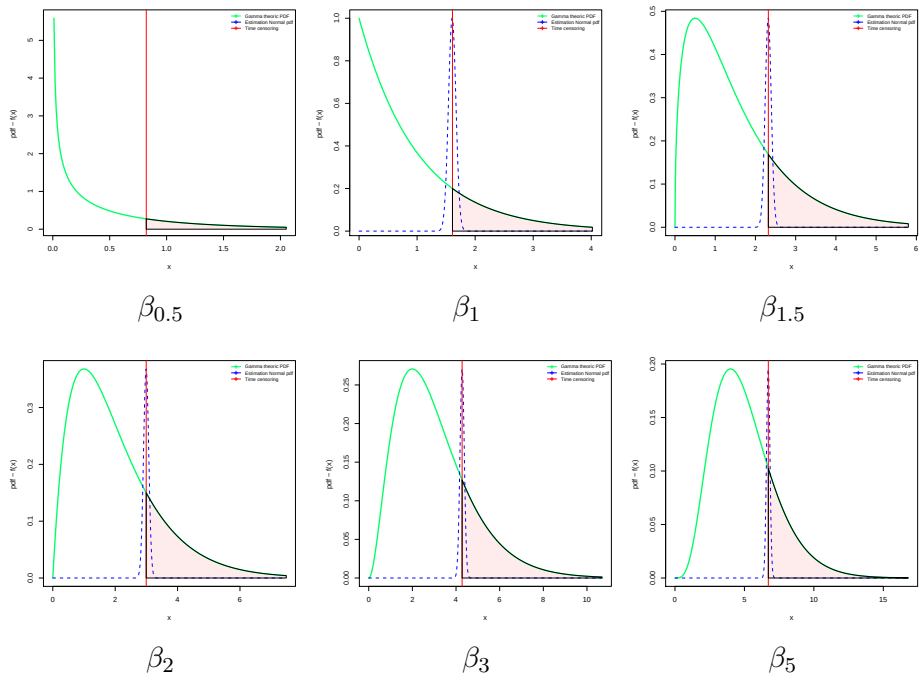


Figure B.4: Simulation right type I, Gamma (α), $C = 20\%$, $n = 1000$

B.3 Figures from Log-normal distribution

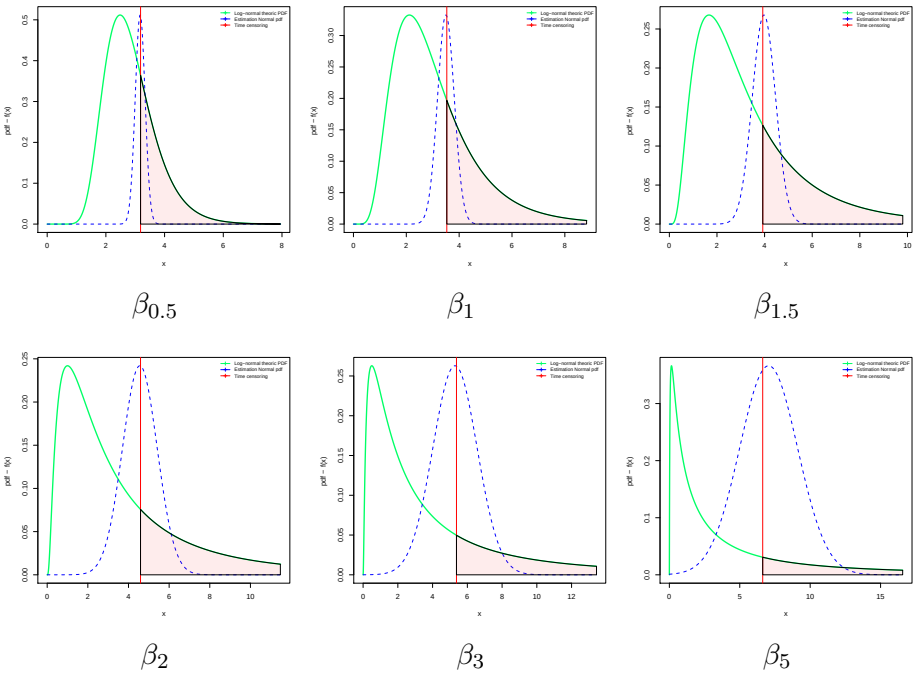


Figure B.5: Simulation right type I, Log-Normal (σ), $C = 30\%$, $n = 50$

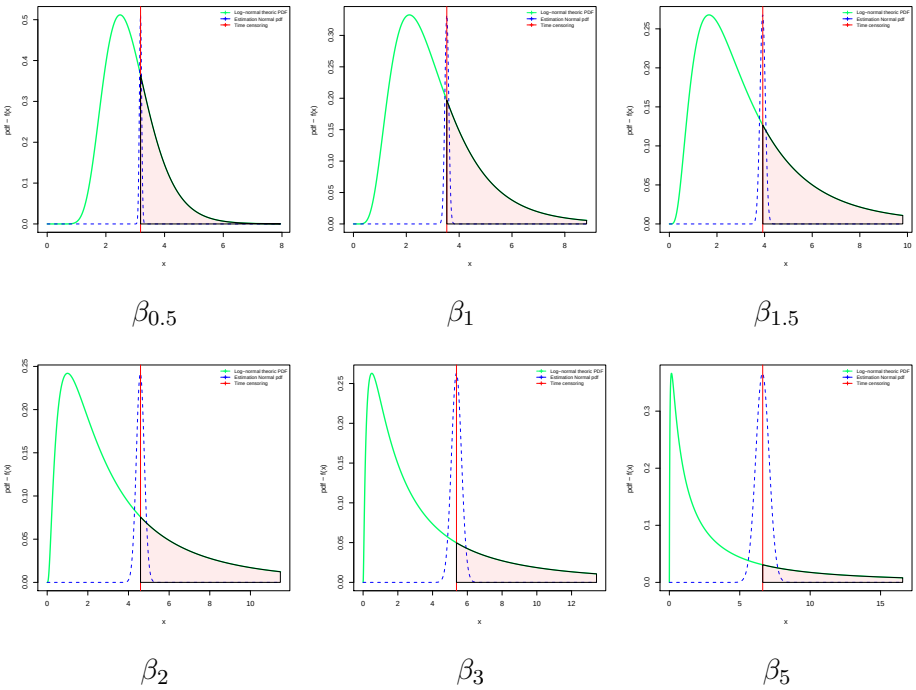


Figure B.6: Simulation right type I, Log-Normal (σ), $C = 30\%$, $n = 1000$

B.4 Figures from Exponential distribution

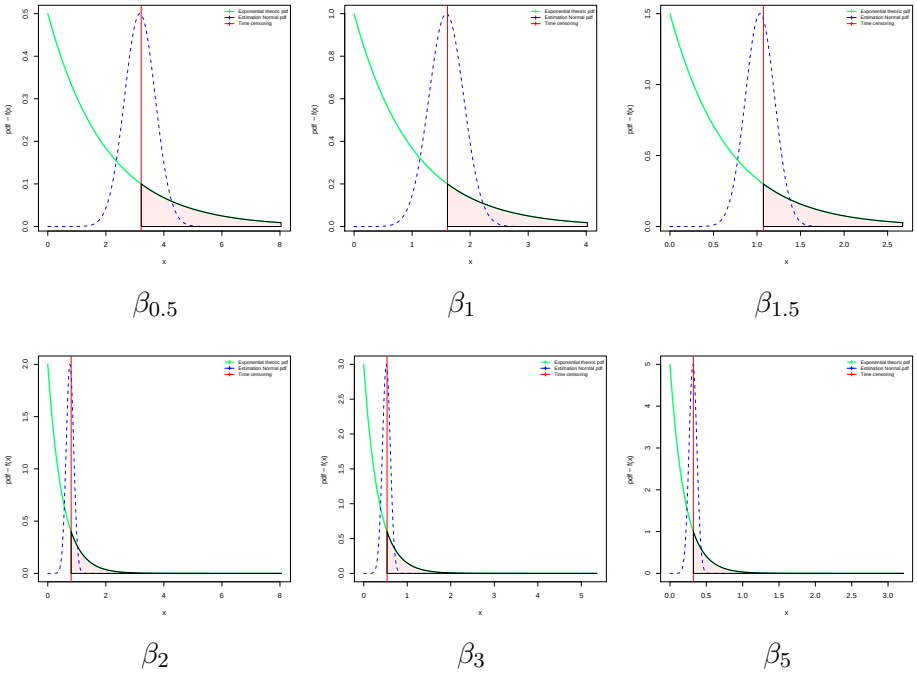


Figure B.7: Simulation right type I, Exponential (λ), $C = 20\%$, $n = 50$

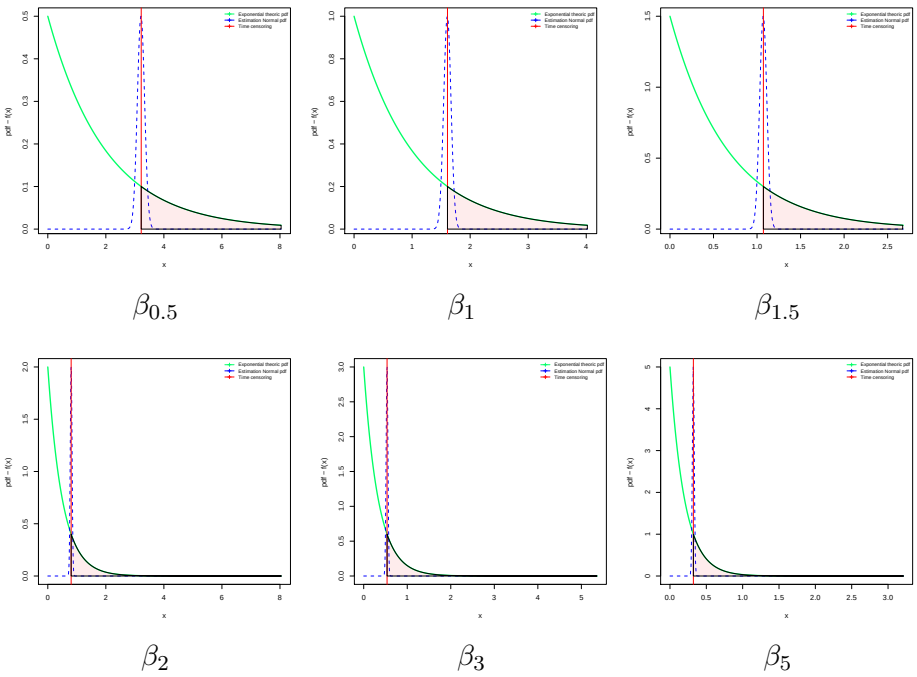
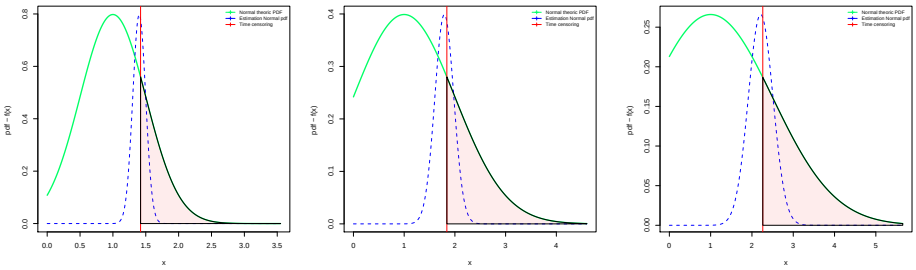
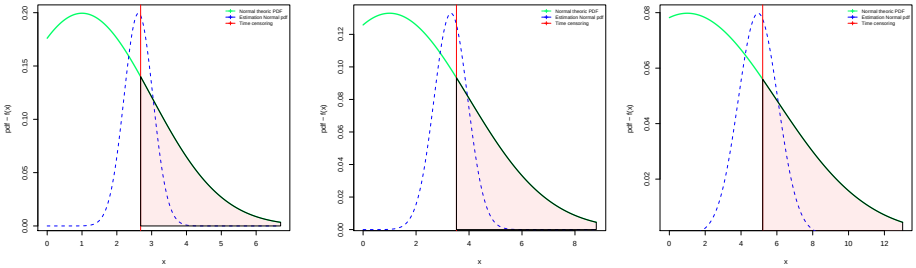


Figure B.8: Simulation right type I, Exponential (λ), $C = 20\%$, $n = 1000$

B.5 Figures from Normal distribution

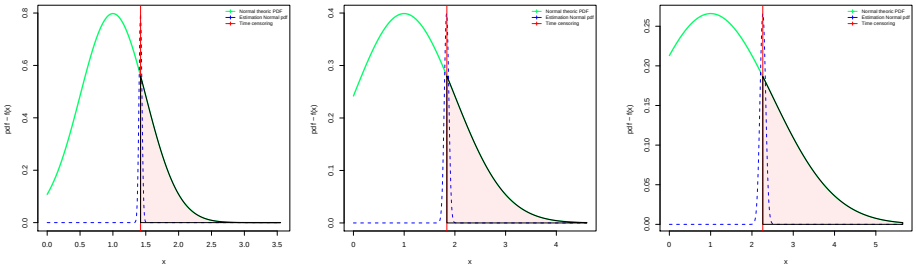


$\beta_{0.5}$ β_1 $\beta_{1.5}$

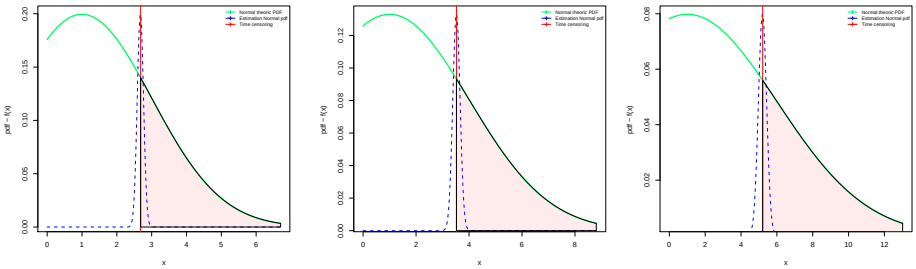


β_2 β_3 β_5

Figure B.9: Simulation right type I, Normal (σ), $C = 20\%$, $n = 50$



$\beta_{0.5}$ β_1 $\beta_{1.5}$



β_2 β_3 β_5

Figure B.10: Simulation right type I, Normal (σ), $C = 20\%$, $n = 1000$

Appendix C

Simulation and test of data censored right fixed type-I

C.1 Weibull distribution

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\beta_{0.5}$	0.97	0	0	0.98	0.96	0.96	0.91	0.92	0.97	0.96	0.95	0.61	0.5	0.95	0.96	0.95	0.07	0.02	0.97	0.95
β_1	0.97	0	0	0.98	0.96	0.95	0.93	0.92	0.95	0.95	0.95	0.6	0.52	0.94	0.94	0.95	0.1	0.02	0.96	0.95
$\beta_{1.5}$	0.96	0	0	0.98	0.95	0.96	0.92	0.93	0.96	0.95	0.95	0.62	0.52	0.96	0.94	0.94	0.09	0	0.95	0.95
β_2	0.97	0	0	0.98	0.94	0.95	0.93	0.92	0.96	0.95	0.96	0.63	0.51	0.96	0.96	0.96	0.09	0.02	0.95	0.95
β_3	0.97	0	0	0.98	0.95	0.95	0.93	0.92	0.96	0.96	0.96	0.63	0.54	0.96	0.96	0.95	0.1	0.01	0.96	0.95
β_5	0.96	0	0	0.96	0.93	0.94	0.92	0.91	0.95	0.94	0.95	0.63	0.54	0.95	0.94	0.96	0.09	0.02	0.96	0.95

Table C.1: Test RNG, fixed right type I, Weibull $(\beta, C\%)$, $\alpha = 0.05$, $\eta = 1$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\beta_{0.5}$	0.93	0	0	0.95	0.89	0.91	0.88	0.87	0.9	0.9	0.91	0.49	0.39	0.91	0.9	0.9	0.05	0	0.89	0.88
β_1	0.94	0	0	0.96	0.91	0.9	0.87	0.87	0.89	0.89	0.92	0.5	0.39	0.9	0.9	0.9	0.04	0.01	0.91	0.89
$\beta_{1.5}$	0.95	0	0	0.96	0.92	0.89	0.85	0.83	0.91	0.91	0.92	0.48	0.39	0.91	0.88	0.92	0.03	0.01	0.92	0.91
β_2	0.93	0	0	0.94	0.89	0.9	0.87	0.86	0.92	0.9	0.91	0.5	0.39	0.91	0.9	0.92	0.05	0.01	0.88	0.88
β_3	0.93	0	0	0.95	0.91	0.92	0.86	0.84	0.9	0.91	0.91	0.53	0.42	0.92	0.9	0.9	0.04	0	0.91	0.92
β_5	0.95	0	0	0.95	0.9	0.91	0.84	0.86	0.9	0.9	0.9	0.49	0.38	0.9	0.91	0.91	0.05	0.01	0.92	0.9

Table C.2: Test RNG, fixed right type I, Weibull $(\beta, C\%)$, $\alpha = 0.01$, $\eta = 1$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
N	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
50	1.9	1.4	0.4	1.9	1.8	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.8	1.7	1.9	1.9
100	0.95	0.42	0.05	0.97	0.93	0.94	0.97	0.94	0.99	0.98	0.95	0.94	0.88	0.97	0.97	0.96	0.88	0.66	0.98	0.92
500	0.2	0	0	0.2	0.19	0.19	0.18	0.19	0.2	0.2	0.2	0.17	0.15	0.19	0.19	0.18	0.08	0.03	0.18	0.19
1000	0.1	0	0	0.1	0.1	0.1	0.09	0.09	0.1	0.09	0.1	0.06	0.06	0.1	0.1	0.1	0.01	0	0.1	0.1
5000	0.02	0	0	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0.02	0.02	0.02	0	0	0.02	0.02
10000	0.01	0	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0	0.01	0.01	0.01	0	0	0.01	0.01

Table C.3: Test RNG, fixed right type I, Weibull $(n, C\%)$, $\alpha = 0.05$, $\eta = 1$, $\beta = 1$

C.2 Gamma distribution

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\alpha_{0.5}$	0.98	0	0	0.98	0.95	0.96	0.92	0.93	0.95	0.95	0.96	0.63	0.54	0.95	0.95	0.95	0.08	0.01	0.95	0.94
α_{1}	0.97	0	0	0.97	0.94	0.96	0.92	0.92	0.96	0.95	0.94	0.62	0.52	0.96	0.96	0.94	0.09	0.02	0.95	0.95
$\alpha_{1.5}$	0.97	0	0	0.98	0.95	0.95	0.93	0.91	0.96	0.96	0.95	0.64	0.55	0.94	0.96	0.95	0.08	0.01	0.94	0.95
α_{2}	0.98	0	0	0.98	0.95	0.95	0.92	0.92	0.95	0.96	0.95	0.61	0.54	0.95	0.95	0.96	0.11	0.02	0.96	0.94
α_{3}	0.97	0	0	0.98	0.95	0.96	0.92	0.93	0.96	0.95	0.96	0.63	0.48	0.95	0.96	0.96	0.08	0.01	0.96	0.96
α_{5}	0.97	0	0	0.97	0.94	0.95	0.92	0.94	0.94	0.96	0.96	0.63	0.52	0.95	0.96	0.96	0.09	0.01	0.96	0.95

Table C.4: Test RNG, fixed right type I, Gamma $(\alpha, C\%)$, $\alpha = 0.05$, $\lambda = 10$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.10}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\alpha_{0.5}$	0.93	0	0	0.95	0.91	0.9	0.85	0.86	0.93	0.9	0.89	0.52	0.41	0.9	0.91	0.92	0.06	0.01	0.92	0.91
α_{1}	0.94	0	0	0.96	0.89	0.9	0.86	0.85	0.89	0.88	0.91	0.54	0.4	0.9	0.9	0.91	0.05	0.01	0.91	0.91
$\alpha_{1.5}$	0.94	0	0	0.96	0.9	0.9	0.87	0.88	0.9	0.9	0.9	0.49	0.42	0.91	0.91	0.92	0.04	0.01	0.92	0.89
α_{2}	0.93	0	0	0.95	0.9	0.9	0.86	0.86	0.91	0.9	0.89	0.51	0.42	0.92	0.91	0.91	0.05	0	0.92	0.89
α_{3}	0.94	0	0	0.94	0.9	0.9	0.87	0.86	0.93	0.91	0.9	0.49	0.4	0.91	0.91	0.9	0.06	0.01	0.91	0.9
α_{5}	0.93	0	0	0.95	0.89	0.9	0.88	0.85	0.9	0.9	0.9	0.5	0.43	0.89	0.9	0.91	0.06	0	0.89	0.88

Table C.5: Test RNG, fixed right type I, Gamma $(\alpha, C\%)$, $\alpha = 0.01$, $\lambda = 10$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
N	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
50	2	1.2	0.62	2	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	2	1.9	1.9	1.7	1.6	1.9	1.9
100	0.97	0.39	0.04	1	0.96	0.97	0.93	0.91	0.98	0.95	0.96	0.96	0.89	1	0.99	0.94	0.84	0.77	0.96	0.94
500	0.18	0	0	0.2	0.19	0.19	0.19	0.19	0.18	0.19	0.18	0.15	0.15	0.18	0.19	0.19	0.08	0.03	0.19	0.19
1000	0.1	0	0	0.1	0.1	0.1	0.09	0.09	0.1	0.1	0.1	0.05	0.06	0.1	0.09	0.1	0.01	0	0.09	0.09
5000	0.02	0	0	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0.02	0.02	0.02	0	0	0.02	0.02
10000	0.01	0	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0	0.01	0.01	0.01	0	0	0.01	0.01

Table C.6: Test RNG, fixed right type I, Gamma $(n, C\%)$, $\alpha_l = 0.05$, $\lambda = 10$, $\alpha = 1$

C.3 Normal distribution

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\sigma_{0.5}$	0	0	0	0.98	0.94	0	0.92	0.92	0.95	0.94	0	0.62	0.55	0.96	0.96	0	0.09	0.02	0.94	0.94
σ_1	0	0	0	0.98	0.94	0	0.93	0.92	0.96	0.95	0	0.66	0.55	0.97	0.96	0	0.08	0.01	0.97	0.95
$\sigma_{1.5}$	0	0	0	0.97	0.94	0	0.92	0.92	0.95	0.94	0	0.63	0.5	0.95	0.96	0	0.09	0.02	0.95	0.96
σ_2	0	0	0	0.97	0.95	0	0.93	0.93	0.94	0.94	0	0.64	0.49	0.94	0.95	0	0.1	0.01	0.94	0.95
σ_3	0	0	0	0.97	0.95	0	0.94	0.92	0.96	0.96	0	0.61	0.51	0.95	0.95	0	0.08	0.02	0.95	0.96
σ_5	0	0	0	0.98	0.94	0	0.94	0.93	0.96	0.94	0	0.65	0.54	0.96	0.95	0	0.1	0.01	0.96	0.93

Table C.7: Test RNG, fixed right type I, Normal $(\sigma, C\%)$, $\alpha = 0.05$, $\mu = 1$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.10}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\sigma_{0.5}$	0	0	0	0.96	0.9	0	0.86	0.87	0.9	0.9	0	0.52	0.39	0.9	0.91	0	0.05	0.01	0.92	0.92
σ_1	0	0	0	0.94	0.9	0	0.86	0.86	0.91	0.89	0	0.51	0.4	0.9	0.9	0	0.04	0.01	0.91	0.91
$\sigma_{1.5}$	0	0	0	0.94	0.91	0	0.86	0.86	0.92	0.9	0	0.49	0.39	0.92	0.89	0	0.05	0	0.9	0.9
σ_2	0	0	0	0.95	0.91	0	0.87	0.85	0.9	0.9	0	0.48	0.42	0.91	0.9	0	0.05	0.01	0.9	0.9
σ_3	0	0	0	0.96	0.91	0	0.86	0.87	0.91	0.9	0	0.5	0.41	0.92	0.9	0	0.05	0	0.91	0.9
σ_5	0	0	0	0.97	0.91	0	0.87	0.86	0.89	0.9	0	0.49	0.4	0.91	0.91	0	0.04	0.01	0.91	0.9

Table C.8: Test RNG, fixed right type I, Normal $(\sigma, C\%)$, $\alpha = 0.01$, $\mu = 1$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
N	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
50	1	1.3	0.58	1.9	1.9	0.96	1.9	1.9	1.9	1.9	0.98	1.9	1.9	2	1.9	1.1	1.9	1.6	2	2
100	0.19	0.43	0.07	1	0.96	0.21	0.93	0.96	0.99	0.96	0.14	0.96	0.92	0.94	0.98	0.18	0.8	0.75	0.97	0.94
500	0	0	0	0.2	0.2	0	0.19	0.18	0.19	0.2	0	0.15	0.15	0.19	0.19	0	0.08	0.03	0.18	0.18
1000	0	0	0	0.1	0.1	0	0.09	0.09	0.09	0.09	0	0.06	0.06	0.1	0.1	0	0.01	0	0.09	0.09
5000	0	0	0	0.02	0.02	0	0.02	0.02	0.02	0.02	0	0	0	0.02	0.02	0	0	0	0.02	0.02
10000	0	0	0	0.01	0.01	0	0.01	0.01	0.01	0.01	0	0	0	0.01	0.01	0	0	0	0.01	0.01

Table C.9: Test RNG, fixed right type I, Normal $(n, C\%)$, $\alpha = 0.05$, $\mu = 1$, $\sigma = 1$

C.4 Log-normal distribution

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\sigma_{0.3}$	0.96	0	0	0.98	0.94	0.94	0.92	0.93	0.94	0.94	0.96	0.6	0.54	0.94	0.94	0.95	0.09	0.02	0.95	0.95
$\sigma_{0.5}$	0.98	0	0	0.97	0.94	0.95	0.93	0.93	0.96	0.95	0.94	0.62	0.52	0.96	0.95	0.96	0.09	0.02	0.94	0.94
$\sigma_{0.7}$	0.98	0	0	0.97	0.95	0.95	0.92	0.92	0.94	0.93	0.95	0.64	0.55	0.96	0.94	0.96	0.09	0.02	0.95	0.95
σ_{1}	0.96	0	0	0.98	0.96	0.96	0.93	0.92	0.95	0.95	0.97	0.63	0.5	0.96	0.95	0.96	0.1	0.02	0.96	0.95
$\sigma_{1.3}$	0.96	0	0	0.98	0.95	0.95	0.93	0.92	0.95	0.95	0.95	0.62	0.52	0.94	0.94	0.95	0.09	0.02	0.94	0.94
$\sigma_{1.7}$	0.97	0	0	0.98	0.96	0.96	0.93	0.93	0.96	0.96	0.96	0.63	0.52	0.96	0.95	0.95	0.1	0.02	0.95	0.94

Table C.10: Test RNG, fixed right type I, Log-Nor. $(\sigma, C\%)$, $\alpha = 0.05$, $\mu = 10$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
N	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
50	1.9	1.3	0.44	2	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.9	1.8	2	1.8	1.6	2	1.9
100	1	0.31	0.07	0.97	0.95	0.93	0.96	0.96	0.99	0.98	0.97	0.92	0.9	0.97	0.98	0.94	0.87	0.72	0.94	0.95
500	0.2	0	0	0.2	0.19	0.19	0.18	0.19	0.18	0.18	0.18	0.15	0.15	0.2	0.19	0.19	0.08	0.04	0.2	0.19
1000	0.1	0	0	0.1	0.1	0.1	0.09	0.09	0.1	0.1	0.09	0.07	0.05	0.1	0.1	0.1	0.01	0	0.1	0.1
5000	0.02	0	0	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0	0	0.02	0.02	0.02	0	0	0.02	0.02
10000	0.01	0	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0	0.01	0.01	0.01	0	0	0.01	0.01

Table C.11: Test RNG, fixed right type I, Log-Nor. $(\sigma, C\%)$, $\alpha = 0.01$, $\mu = 10$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.10}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\sigma_{0.3}$	0.92	0	0	0.94	0.89	0.9	0.87	0.85	0.91	0.88	0.9	0.51	0.41	0.9	0.89	0.9	0.04	0	0.9	0.89
$\sigma_{0.5}$	0.93	0	0	0.96	0.91	0.9	0.86	0.85	0.9	0.91	0.9	0.5	0.38	0.92	0.91	0.91	0.05	0.01	0.9	0.9
$\sigma_{0.7}$	0.94	0	0	0.96	0.9	0.9	0.87	0.87	0.91	0.92	0.92	0.5	0.4	0.91	0.88	0.92	0.05	0.01	0.91	0.89
σ_{1}	0.94	0	0	0.95	0.9	0.91	0.86	0.88	0.89	0.89	0.88	0.5	0.42	0.92	0.89	0.92	0.05	0.01	0.91	0.9
$\sigma_{1.3}$	0.95	0	0	0.95	0.89	0.91	0.86	0.87	0.91	0.9	0.92	0.52	0.4	0.9	0.89	0.92	0.05	0.01	0.91	0.91
$\sigma_{1.7}$	0.94	0	0	0.96	0.91	0.91	0.87	0.86	0.92	0.91	0.91	0.53	0.39	0.92	0.9	0.9	0.04	0.01	0.92	0.9

Table C.12: Test RNG, fixed right type I, Log-Nor. $(n, C\%)$, $\alpha = 0.05$, $\mu = 10$, $\sigma = 10$

C.5 Exponential distribution

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\lambda_{0.5}$	0	0	0	0.98	0.96	0	0.91	0.92	0.96	0.95	0	0.62	0.52	0.96	0.96	0	0.09	0.01	0.95	0.95
λ_1	0	0	0	0.98	0.95	0	0.92	0.93	0.95	0.95	0	0.64	0.55	0.96	0.96	0	0.08	0.01	0.96	0.95
$\lambda_{1.5}$	0	0	0	0.98	0.94	0	0.93	0.91	0.95	0.95	0	0.6	0.52	0.94	0.95	0	0.09	0.01	0.96	0.96
λ_2	0	0	0	0.98	0.95	0	0.93	0.93	0.94	0.95	0	0.65	0.5	0.94	0.96	0	0.08	0.01	0.94	0.95
λ_3	0	0	0	0.98	0.94	0	0.93	0.92	0.95	0.95	0	0.64	0.54	0.96	0.94	0	0.09	0.02	0.95	0.94
λ_5	0	0	0	0.97	0.94	0	0.93	0.94	0.94	0.95	0	0.65	0.5	0.96	0.95	0	0.1	0.02	0.96	0.97

Table C.13: Test RNG, fixed right type I, Exponential $(\lambda, C\%)$, $\alpha = 0.05$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.10}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\lambda_{0.5}$	0	0	0	0.95	0.9	0	0.87	0.85	0.92	0.9	0	0.52	0.37	0.9	0.9	0	0.05	0.01	0.91	0.9
λ_1	0	0	0	0.95	0.9	0	0.89	0.86	0.89	0.89	0	0.48	0.4	0.9	0.89	0	0.04	0.01	0.91	0.9
$\lambda_{1.5}$	0	0	0	0.95	0.89	0	0.86	0.86	0.92	0.9	0	0.53	0.41	0.91	0.9	0	0.05	0	0.91	0.9
λ_2	0	0	0	0.95	0.89	0	0.86	0.84	0.9	0.9	0	0.47	0.42	0.9	0.91	0	0.04	0	0.92	0.9
λ_3	0	0	0	0.97	0.91	0	0.86	0.87	0.9	0.9	0	0.5	0.42	0.9	0.89	0	0.05	0.01	0.9	0.9
λ_5	0	0	0	0.95	0.91	0	0.85	0.9	0.9	0.89	0	0.5	0.42	0.92	0.91	0	0.04	0.01	0.91	0.92

Table C.14: Test RNG, fixed right type I, Exponential $(\lambda, C\%)$, $\alpha = 0.01$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
N	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
50	0	1.3	0.52	2	1.9	1.9	1.9	1.9	1.9	1.9	0	1.9	1.9	1.9	1.9	0	1.8	1.7	2	1.9
100	0	0.35	0.03	0.98	0.94	0.95	0.96	0.94	0.98	0.97	0	0.95	0.93	0.97	0.9	0	0.8	0.74	0.95	0.95
500	0	0	0	0.2	0.19	0	0.19	0.19	0.19	0.19	0	0.15	0.15	0.19	0.19	0	0.08	0.04	0.19	0.19
1000	0	0	0	0.1	0.09	0	0.09	0.08	0.1	0.1	0	0.06	0.04	0.09	0.1	0	0.01	0	0.09	0.1
5000	0	0	0	0.02	0.02	0	0.02	0.02	0.02	0.02	0	0	0	0.02	0.02	0	0	0	0.02	0.02
10000	0	0	0	0.01	0.01	0	0.01	0.01	0.01	0.01	0	0	0	0.01	0.01	0	0	0	0.01	0.01

Table C.15: Test RNG, fixed right type I, Exponential $(n, C\%)$, $\alpha = 0.05$, $\lambda = 1$

C.6 Global results

		$C_{(5\%)}$	$C_{(10\%)}$	$C_{(20\%)}$	$C_{(30\%)}$
Weibull	$\alpha_{0.05}$	Adm	OK	Adm	Adm
	$\alpha_{0.1}$	NOK	Adm	NOK	NOK
normal	$\alpha_{0.05}$	NOK	OK	NOK	NOK
	$\alpha_{0.1}$	NOK	NOK	NOK	NOK
lognormal	$\alpha_{0.05}$	Adm	OK	Adm	Adm
	$\alpha_{0.1}$	NOK	Adm	NOK	NOK
gamma	$\alpha_{0.05}$	Adm	OK	Adm	Adm
	$\alpha_{0.1}$	NOK	Adm	NOK	NOK
exponencial	$\alpha_{0.05}$	NOK	OK	NOK	NOK
	$\alpha_{0.1}$	NOK	NOK	NOK	NOK

Table C.16: Global Results of Simulation Test - Right Fixed Type I - Shape Factor

		$C_{(5\%)}$	$C_{(10\%)}$	$C_{(20\%)}$	$C_{(30\%)}$
Weibull	$\alpha_{0.05}$	NOK	NOK	NOK	NOK
	$\alpha_{0.1}$	NOK	NOK	NOK	NOK
normal	$\alpha_{0.05}$	NOK	NOK	NOK	NOK
	$\alpha_{0.1}$	NOK	NOK	NOK	NOK
lognormal	$\alpha_{0.05}$	NOK	NOK	NOK	NOK
	$\alpha_{0.1}$	NOK	NOK	NOK	NOK
gamma	$\alpha_{0.05}$	NOK	NOK	NOK	NOK
	$\alpha_{0.1}$	NOK	NOK	NOK	NOK
exponencial	$\alpha_{0.05}$	NOK	NOK	NOK	NOK
	$\alpha_{0.1}$	NOK	NOK	NOK	NOK

Table C.17: Global Results of Simulation Test - Right Fixed Type I - Sample n

		$C_{(5\%)}$	$C_{(10\%)}$	$C_{(20\%)}$	$C_{(30\%)}$
Weibull	$\alpha_{0.05}$	11.15	12.58	12.3	11.98
	$\alpha_{0.1}$	11.03	12.48	12.35	11.98
normal	$\alpha_{0.05}$	10.88	12.08	11.92	11.63
	$\alpha_{0.1}$	10.78	12.08	11.9	11.63
lognormal	$\alpha_{0.05}$	10.93	12.25	12.05	11.75
	$\alpha_{0.1}$	10.83	12.25	12.05	11.75
gamma	$\alpha_{0.05}$	10.85	12.12	11.95	11.67
	$\alpha_{0.1}$	10.82	12.13	11.95	11.67
exponencial	$\alpha_{0.05}$	11.15	12.58	12.3	11.98
	$\alpha_{0.1}$	11.03	12.48	12.35	11.98

Table C.18: Global table of time in simulation to different shape factors (minutes)

		$C_{(5\%)}$	$C_{(10\%)}$	$C_{(20\%)}$	$C_{(30\%)}$
Weibull	$\alpha_{0.05}$	13.58	14.53	14.4	14.18
	$\alpha_{0.1}$	13.57	14.65	14.5	14.28
normal	$\alpha_{0.05}$	13.73	14.77	14.57	14.37
	$\alpha_{0.1}$	13.68	14.68	14.55	14.37
lognormal	$\alpha_{0.05}$	13.43	14.42	14.23	14
	$\alpha_{0.1}$	13.28	14.37	14.18	14.03
gamma	$\alpha_{0.05}$	13.5	14.47	14.35	14.13
	$\alpha_{0.1}$	13.45	14.48	14.32	14.12
exponencial	$\alpha_{0.05}$	13.73	14.71	14.55	14.41
	$\alpha_{0.1}$	13.64	14.72	14.56	0.01

Table C.19: Global table of time in simulation to different number of samples (minutes)

Appendix D

Simulation and test of data censored right type-II

D.1 Weibull distribution

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\beta_{0.5}$	1	0.1	0	0.09	0	1	0.96	0.85	0.96	0.94	1	0.94	0.17	0.94	0.9	1	0.83	0	0.89	0.56
β_1	1	0.08	0	0.08	0	1	0.95	0.87	0.95	0.94	1	0.95	0.16	0.94	0.9	1	0.83	0	0.89	0.56
$\beta_{1.5}$	1	0.07	0	0.09	0	1	0.95	0.86	0.95	0.95	1	0.95	0.17	0.95	0.89	1	0.85	0	0.88	0.56
β_2	1	0.09	0	0.08	0	1	0.95	0.86	0.95	0.96	1	0.95	0.16	0.94	0.88	1	0.83	0	0.89	0.54
β_3	1	0.08	0	0.08	0	1	0.96	0.86	0.96	0.96	1	0.93	0.16	0.96	0.89	1	0.83	0	0.9	0.58
β_5	1	0.07	0	0.08	0	1	0.96	0.86	0.96	0.95	1	0.94	0.18	0.95	0.9	1	0.85	0	0.88	0.56

Table D.1: Test RNG, right type II, Weibull $(\beta, C\%)$, $\alpha = 0.05$, $\eta = 1$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\beta_{0.5}$	1	0.04	0	0.06	0	1	0.88	0.78	0.9	0.9	1	0.88	0.11	0.89	0.8	1	0.73	0	0.84	0.47
β_1	1	0.04	0	0.05	0	1	0.91	0.76	0.9	0.88	1	0.89	0.1	0.9	0.82	1	0.76	0	0.84	0.45
$\beta_{1.5}$	1	0.04	0	0.05	0	1	0.89	0.78	0.92	0.9	1	0.89	0.1	0.91	0.81	1	0.76	0	0.82	0.44
β_2	1	0.03	0	0.04	0	1	0.91	0.8	0.93	0.92	1	0.88	0.1	0.9	0.83	1	0.77	0	0.8	0.42
β_3	1	0.03	0	0.05	0	1	0.89	0.78	0.9	0.9	1	0.91	0.1	0.91	0.85	1	0.76	0	0.82	0.45
β_5	1	0.04	0	0.05	0	1	0.9	0.76	0.89	0.89	1	0.9	0.11	0.9	0.82	1	0.76	0	0.8	0.43

Table D.2: Test RNG, right type II, Weibull $(\beta, C\%)$, $\alpha = 0.01$, $\eta = 1$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
N	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
50	2	1.9	1.9	2	1.9	2	1.9	1.9	2	1.9	2	1.9	1.9	1.9	1.8	2	1.9	1.9	1.9	1.9
100	1	0.91	0.94	0.96	0.93	1	0.98	0.98	0.98	0.96	1	0.92	0.93	0.96	0.95	1	0.88	0.93	0.93	0.91
500	0.2	0.19	0.19	0.19	0.19	0.2	0.2	0.19	0.19	0.2	0.2	0.19	0.19	0.19	0.19	0.2	0.19	0.19	0.19	0.19
1000	0.1	0.09	0.1	0.09	0.09	0.1	0.09	0.09	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.09	0.1	0.09	0.09
5000	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
10000	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Table D.3: Test RNG, right type II, Weibull $(n, C\%)$, $\alpha = 0.05$, $\eta = 1$, $\beta = 1$

D.2 Gamma distribution

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.05}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\alpha_{0.5}$	1	0.08	0	0.08	0	1	0.94	0.88	0.95	0.95	1	0.95	0.18	0.96	0.9	1	0.83	0	0.88	0.57
α_1	1	0.09	0	0.08	0	1	0.95	0.86	0.95	0.96	1	0.94	0.16	0.95	0.89	1	0.84	0	0.88	0.55
$\alpha_{1.5}$	1	0.07	0	0.09	0	1	0.95	0.87	0.95	0.95	1	0.93	0.18	0.94	0.9	1	0.84	0	0.9	0.55
α_2	1	0.09	0	0.08	0	1	0.95	0.88	0.95	0.94	1	0.94	0.17	0.94	0.9	1	0.84	0	0.9	0.56
α_3	1	0.08	0	0.08	0	1	0.94	0.87	0.95	0.95	1	0.94	0.18	0.95	0.89	1	0.85	0	0.9	0.54
α_5	1	0.09	0	0.08	0	1	0.96	0.86	0.95	0.95	1	0.95	0.17	0.95	0.9	1	0.85	0	0.89	0.56

Table D.4: Test RNG, right type II, Gamma $(\alpha, C\%)$, $\alpha = 0.05$, $\lambda = 1$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
$\alpha_{0.10}$	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
$\alpha_{0.5}$	1	0.04	0	0.04	0	1	0.91	0.77	0.91	0.9	1	0.88	0.12	0.89	0.82	1	0.75	0	0.81	0.44
α_1	1	0.04	0	0.05	0	1	0.92	0.76	0.91	0.9	1	0.89	0.09	0.9	0.82	1	0.78	0	0.78	0.4
$\alpha_{1.5}$	1	0.04	0	0.06	0	1	0.9	0.75	0.93	0.91	1	0.88	0.08	0.9	0.83	1	0.75	0	0.81	0.44
α_2	1	0.04	0	0.05	0	1	0.89	0.77	0.91	0.9	1	0.88	0.1	0.89	0.8	1	0.75	0	0.82	0.44
α_3	1	0.03	0	0.05	0	1	0.9	0.78	0.89	0.89	1	0.88	0.1	0.9	0.83	1	0.75	0	0.81	0.45
α_5	1	0.03	0	0.06	0	1	0.92	0.75	0.9	0.89	1	0.88	0.1	0.89	0.83	1	0.74	0	0.8	0.45

Table D.5: Test RNG, right type II, Gamma $(\alpha, C\%)$, $\alpha = 0.01$, $\lambda = 1$, $n = 100$

	$\mathbf{C}_{5\%}$					$\mathbf{C}_{10\%}$					$\mathbf{C}_{20\%}$					$\mathbf{C}_{30\%}$				
N	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s	R_u	R_a	T_u	B_a	K_s
50	2	2	1.8	1.9	1.8	2	1.9	1.9	2	1.9	2	1.9	1.9	1.9	1.9	2	1.9	1.8	1.9	1.9
100	1	0.96	0.93	0.91	0.98	1	0.97	0.92	0.94	0.94	1	0.94	0.94	0.92	0.91	1	0.93	0.96	0.98	0.95
500	0.2	0.19	0.19	0.19	0.19	0.2	0.18	0.19	0.19	0.19	0.2	0.19	0.19	0.2	0.2	0.2	0.18	0.19	0.19	0.2
1000	0.1	0.09	0.09	0.09	0.1	0.1	0.1	0.09	0.1	0.1	0.1	0.09	0.1	0.09	0.09	0.1	0.09	0.09	0.09	0.1
5000	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
10000	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Table D.6: Test RNG, right type II, Gamma $(n, C\%)$, $\alpha = 0.05$, $\lambda = 1$, $\alpha = 1$

D.3 Global results

The final result for the variation of distribution parameters is shown in the following table:

		$C_{(5\%)}$	$C_{(10\%)}$	$C_{(20\%)}$	$C_{(30\%)}$
	$\alpha_{0.05}$	OK	OK	OK	OK
weibull	$\alpha_{0.1}$	OK	OK	OK	OK
	$\alpha_{0.05}$	OK	OK	OK	OK
gamma	$\alpha_{0.1}$	OK	OK	OK	OK

Table D.7: Global Results of Simulation Test - Right Type II - Shape Factor

		$C_{(5\%)}$	$C_{(10\%)}$	$C_{(20\%)}$	$C_{(30\%)}$
	$\alpha_{0.05}$	NOK	NOK	NOK	NOK
weibull	$\alpha_{0.1}$	NOK	NOK	NOK	NOK
	$\alpha_{0.05}$	NOK	NOK	NOK	NOK
gamma	$\alpha_{0.1}$	NOK	NOK	NOK	NOK

Table D.8: Global Results of Simulation Test - Right Type II - Sample n

And in the study also possible to measure the simulation times, with a computer with a processor Pentium 4, 8GB of ram.

		$C_{(5\%)}$	$C_{(10\%)}$	$C_{(20\%)}$	$C_{(30\%)}$
	$\alpha_{0.05}$	0	0	0	0
weibull	$\alpha_{0.1}$	0	0	0	0
	$\alpha_{0.05}$	0.02	0.02	0.02	0.02
gamma	$\alpha_{0.1}$	0.02	0.02	0.02	0.02

Table D.9: Global table of time in simulation to different shape factors (minutes)

		$C_{(5\%)}$	$C_{(10\%)}$	$C_{(20\%)}$	$C_{(30\%)}$
	$\alpha_{0.05}$	0.38	1.07	0.85	0.68
weibull	$\alpha_{0.1}$	0.38	1.07	0.87	0.68
	$\alpha_{0.05}$	0.38	1.07	0.87	0.7
gamma	$\alpha_{0.1}$	0.38	1.08	0.88	0.68

Table D.10: Global table of time in simulation to different number of samples (minutes)

References

Aydin, D. and Yilmaz, E. (2018). Modified estimators in semiparametric regression models with right-censored data. *Journal of Statistical Computation and Simulation*, 88(8):1470–1498.

Balakrishnan, N., Balakrishnan, N., and Aggarwala, R. (2000). *Progressive censoring: theory, methods, and applications*. Springer Science & Business Media.

Balakrishnan, N. and Kannan, N. (2001). Ch. 14 point and interval estimation for parameters of the logistic distribution based on progressively type-ii censored samples. *Handbook of statistics*, 20:431–456.

Balakrishnan, N. and Mitra, D. (2012). Left truncated and right censored weibull data and likelihood inference with an illustration. *Computational Statistics & Data Analysis*, 56(12):4011–4025.

Banks, J., Carson, J. S., Nelson, B. L., and Nicol, D. M. (2013). *Discrete-event system simulation: Pearson new international edition*. Pearson Higher Ed.

Barlow, R. E. and Proschan, F. (1975). Statistical theory of reliability and life testing: Probability models. Technical report, Florida State Univ Tallahassee.

Birolini, A. (2017). *Reliability Engineering: Theory and Practice*. Springer-Verlag Berlin Heidelberg, 8 edition.

Bloch, H. and Budris, A. (2004). *Pump User's Handbook: Life Extension*. Fairmont Press, New York, NY, USA.

Bollinger, R. and Salvia, A. (1982). Consecutive-k-out-of-n: F networks. *IEEE Transactions on reliability*, 31(1):53–56.

Bradley, E. (2016). *Reliability Engineering: A Life Cycle Approach*. CRC Press, Taylor Francis Group, 2016.

Burton, A., Altman, D. G., Royston, P., and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in medicine*, 25(24):4279–4292.

Caeiro, F. and Mateus, A. (2014). *Randtests: Testing randomness in R*. R package version 1.0.

Checkland, P. and Holwell, S. (1998). *Information, Systems and Information Systems: Making Sense of the Field*. John Wiley & Sons, Inc., New York, NY, USA.

Chiang, D. T. and Niu, S.-C. (1981). Reliability of consecutive-k-out-of-n: F system. *IEEE Transactions on Reliability*, 30(1):87–89.

Childs, A. and Balakrishnan, N. (2000). Conditional inference procedures for the laplace distribution when the observed samples are progressively censored. *Metrika*, 52(3):253–265.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Dixit, J. (2006). *Programming in C and Numerical Analysis*. Firewall Media, New Delhi, 2 edition.

Dubi, A. (2000). *Monte Carlo Applications in Systems Engineering*. Wiley Interscience. Wiley, New York.

Eckhardt, R. (1987). Stan ulam, john von neumann, and the monte carlo method. *Los Alamos Science*, 15(131-136):30.

Escobar, L. A. and Meeker, W. Q. (1999). Statistical prediction based on censored life data. *Technometrics*, 41(2):113–124.

Flynn, D. and Diaz, O. (1996). *Information Modelling: An International Perspective*. Prentice Hall, London.

Freeman, L. J. (2011). A cautionary tale: Small sample size concerns for grouped lifetime data. *Quality Engineering*, 23(2):134–141.

Galp, C. (2018). Refining and logistics, process, accessed 8 de december 2018, https://www.galp.com/corp/en/about-us/our-businesses/refining-and-marketing/sourcing-refining-logistics.

Gaspar, D. (2003). A análise organizacional na especificação dos sistemas de informação em gestão da manutenção. Master's thesis, Faculdade de Engenharia da Universidade do Porto.

Genschel, U. and Meeker, W. Q. (2010). A comparison of maximum likelihood and median-rank regression for weibull estimation. *Quality Engineering*, 22(4):236–255.

Gijbels, I. (2010). Censored data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2):178–188.

Girdhar, P. and Moniz, O. (2005). *Practical centrifugal pumps : design, operation and maintenance*. Practical professional books from Elsevier. Newnes.

Guo, H., Goldsman, D., Tsui, K.-L., Zhou, Y., and Wong, S.-Y. (2016). Using simulation and optimisation to characterise durations of emergency department service times with incomplete data. *International Journal of Production Research*, 54(21):6494–6511.

Gurubilli, P. R. (2010). Random number generation and its better technique. Master's thesis, Thapar University, Patiala, India.

Halabi, S. and Singh, B. (2004). Sample size determination for comparing several survival curves with unequal allocations. *Statistics in Medicine*, 23(11):1793–1815.

Hipel, K., McLeod, A., and Fosu, P. (1986). Empirical power comparisons of some tests for trend. In *Developments in Water Science*, volume 27, pages 347–362. Elsevier.

Horst, R. (2009). *The Weibull Distribution. A Handbook.* Chapman and Hall/CRC Taylor Francis Group.

Jeng, S.-L. and Meeker, W. Q. (2001). Parametric simultaneous confidence bands for cumulative distributions from censored data. *Technometrics*, 43(4):450–461.

Jewell, N., Kimber, A., Lee, M., and Whitmore, G., editors (1996). *Lifetime Data: Models in Reliability and Survival Analysis.* Springer US, New York, NY, USA, 1 edition.

Kamat, S. J. and Riley, M. W. (1975). Determination of reliability using event-based monte carlo simulation. *IEEE transactions on reliability*, 24(1):73–75.

Kang, S.-B. (2003). Approximate mles for exponential distribution under multiple type-ii censoring. *Journal of the Korean Data and Information Science Society*, 14(4):983–988.

Kaufmann, A., Grouchko, D., and Cruon, R. (1977). *Mathematical Models for the Study of the Reliability of Systems.* Mathematics in Science and Engineering : a series of monographs and textbooks. Academic Press.

Kim, C. and Lee, H. K. (1992). A monte carlo simulation algorithm for finding mtbf. *IEEE transactions on reliability*, 41(2):193–195.

Knuth, D. E. (2011). *Art of Computer Programming, Volumes 1-4A Boxed Set.* Addison-Wesley Professional.

Kolmogorov, A. (2013). *Foundations of the Theory of Probability.* Martino Fine Books, reprint of 1956 second edition.

Kontoleon, J. (1980). Reliability determination of a r-successive-out-of-n: F system. *IEEE Transactions on Reliability*, 29(5):437–437.

KSB, C. (2018). Centrifugal pump lexicon - refinery pump, ksb group 2019, accessed 8 de december 2018, https://www.ksb.com/centrifugal-pump-lexicon/refinery-pump/192120/.

L'Ecuyer, P. (2012). Random number generation. In *Handbook of computational statistics*, pages 35–71. Springer.

L'Ecuyer, P. and Simard, R. (2007). Testu01: Ac library for empirical testing of random number generators. *ACM Transactions on Mathematical Software (TOMS)*, 33(4):22.

Lu, L. and Anderson-Cook, C. M. (2017). Choosing a reliability inspection plan for interval censored data. *Quality Engineering*, 29(3):512–535.

Maddison, R. (1983). *Information system methodologies.* Monographs in informatics. Wiley Heyden, on behalf of British Computer Society.

Marsaglia, G., Tsang, W. W., et al. (2002). Some difficult-to-pass tests of randomness. *Journal of Statistical Software*, 7(3):1–9.

Marsaglia, G. and Tsay, L.-H. (1985). Matrices and the structure of random number sequences. *Linear algebra and its applications*, 67:147–156.

Meeker, W. and Escobar, L. (2014). *Statistical Methods for Reliability Data.* Wiley Series in Probability and Statistics. Wiley.

Miloslavsky, M., Keleş, S., van der Laan, M. J., and Butler, S. (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):239–257.

Mitra, D. (2012). *Likelihood inference for left truncated and right censored lifetime data.* PhD thesis, McMaster University, Hamilton, Ontario - USA.

Mousa, M. A. and Al-Sagheer, S. (2005). Bayesian prediction for progressively type-ii censored data from the rayleigh model. *Communications in Statistics—Theory and Methods*, 34(12):2353–2361.

Mousa, M. A. and Jaheen, Z. (2002). Statistical inference for the burr model based on progressively censored data. *Computers & Mathematics with Applications*, 43(10-11):1441–1449.

Muhammad, M., Abd Majid, M. A., and Ibrahim, N. A. (2010). A case study of reliability assessment for centrifugal pumps in a petrochemical plant. In Kiritsis, D., Emmanouilidis, C., Koronios, A., and Mathew, J., editors, *Engineering Asset Lifecycle Management*, pages 398–404, London. Springer London.

Murthy, D. P., Xie, M., and Jiang, R. (2004). *Weibull models*, volume 505. John Wiley & Sons.

Nelson, W. (1985). Weibull analysis of reliability data with few or no failures. *Journal of Quality Technology*, 17(3):140–146.

O'Connor, P. and Kleyner, A. (2012). *Practical reliability engineering.* John Wiley & Sons.

Olteanu, D. and Freeman, L. (2010). The evaluation of median-rank regression and maximum likelihood estimation techniques for a two-parameter weibull distribution. *Quality Engineering*, 22(4):256–272.

Orkand, D. S. (1960). A monte carlo method for determining lower confidence limits for system reliability on the basis of sample component data. Technical report, *Picatinny Arsenal Dover Nj.*

Palgrave, R. (2003). *Troubleshooting Centrifugal Pumps and their systems.* Elsevier Science.

Pampaka, M., Hutcheson, G., and Williams, J. (2016). Handling missing data: analysis of a challenging data set using multiple imputation. *International Journal of Research & Method in Education*, 39(1):19–37.

Park, B. S. (1985). n-person single games and n-component reliability structures.

Rausand, M. and Høyland, A. (2004). *System Reliability Theory: Models, Statistical Methods, and Applications.* Wiley Series in Probability. Wiley.

Rice, R. E. and Moore, A. H. (1983). A monte-carlo technique for estimating lower confidence limits on system reliability using pass-fail data. *IEEE Transactions on Reliability*, 32(4):366–369.

Ross, S. (2012). *Simulation.* Knovel Library. Elsevier Science, 5th edition.

Ross, S. (2017). *Introductory Statistics*. Elsevier Science, 4th edition.

Rubinstein, R. and Kroese, D. (2016). *Simulation and the Monte Carlo Method*. Wiley Series in Probability and Statistics. Wiley.

Sarhan, A. M. and Abuammoh, A. (2008). Statistical inference using progressively type-ii censored data with random scheme. In *International Mathematical Forum*, volume 35, pages 1713–1725.

Shapley, L. S. (1953). *A value for n-person games*, volume 2, pages 307–317. Princeton: Princeton University Press.

Sheikh, A. K., Al-Garni, A. Z., and Affan Badar, M. (1996). Reliability analysis of aeroplane tyres. *International Journal of Quality & Reliability Management*, 13(8):28–38.

Silva, J. L. H. (2016). *Análise de Falhas em Elementos Mecânicos: Aplicação de Métodos Fiabilísticos*. PhD thesis, FEUP, UP.

Sokolowski, J. A. and Banks, C. M. (2010). *Modeling and simulation fundamentals: theoretical underpinnings and practical domains*. John Wiley & Sons.

Soliman, A. A. (2005). Estimation of parameters of life from progressively censored data using burr-xii model. *IEEE Transactions on Reliability*, 54(1):34–42.

Somboonsavatdee, A., Nair, V. N., and Sen, A. (2007). Graphical estimators from probability plots with right-censored data. *Technometrics*, 49(4):420–429.

Team, R. C. (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Vanbrabant, L., Martin, N., Ramaekers, K., and Braekers, K. (2019). Quality of input data in emergency department simulations: Framework and assessment techniques. *Simulation Modelling Practice and Theory*, 91:83–101.

Volterman, W. D. (2011). *On some inferential aspects for Type-II and progressive Type-II censoring*. PhD thesis, McMaster University, Hamilton, Ontario - USA.

Vose, D. (2008). *Risk analysis: a quantitative guide*. John Wiley & Sons.

Wang, P. and Coit, D. W. (2005). Repairable systems reliability trend tests and evaluation. In *Proceedings of the Annual Reliability and Maintainability Symposium*, pages 416–421. IEEE.

Zakaria, R. S. (1989). *The monotonicity of component importance measures in linear consecutive-k-out-of-n systems*. PhD thesis, Iowa State University, Ames, Iowa - USA.

Zhang, C. W., Zhang, T., Xu, D., and Xie, M. (2013). Analyzing highly censored reliability data without exact failure times: an efficient tool for practitioners. *Quality Engineering*, 25(4):392–400.

Zhang, W., Zhang, Y., Chaloner, K., and Stapleton, J. T. (2009). Imputation methods for doubly censored hiv data. *Journal of statistical computation and simulation*, 79(10):1245–1257.

Zhou, X., Feng, Y., and Du, X. (2017). Quantile regression for interval censored data. *Communications in Statistics-Theory and Methods*, 46(8):3848–3863.

Zhou, Y. (2010). *Asset life prediction and maintenance decision-making using a non-linear non-Gaussian state space model*. PhD thesis, Queensland University of Technology.