Matthias Thomas Maurer, BSc

Statistical Analysis of Fatigue Data in Mechanical Engineering

MASTER THESIS

written to obtain the academic degree of a

Master of Science (MSc)

Financial and Actuarial Mathematics

Graz University of Technology

Supervisor:
Univ.-Prof.i.R. Dipl.-Ing. Dr.techn.
Ernst Stadlober

Institute of Statistics

Graz, June, 2017

EIDESSTATTLICHE ERKLÄRUNG

AFFIDAVIT

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit identisch.

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Datum/Date	Unterschrift/Signature

Abstract

This thesis discusses different statistical approaches to fatigue data focused on the total life approach. Different statistical models are introduced, analysed, and used to model an exemplary data set from the literature.

Among these models we find two very common models, the Linear Model and the Generalized Linear Model, representing de facto the status quo. Additionally, we introduce two alternative approaches; In the first one, we are modelling the distribution parameters of the chosen distribution due to the characteristic behaviour of fatigue data (called Parameter Regression Model); The second one is intended to avoid or relax the problem of few data points by standardizing the observations to a common distribution (called Standardizing Approach Model).

A comparison of the introduced models with the free statistic software R in terms of their ability to describe the data's median and scatter leads to a model ranking. In this ranking the Standardizing Approach Model is listed first, followed by the Generalized Linear Model. The next best models are the Parameter Approach Model at the same level as the Linear Model with three predictors. The last listed model is the Linear Model with two predictors. All statistical analyses were carried out using the open source program R (version 3.2.3).

Zusammenfassung

In dieser Masterarbeit werden verschiedene statistische Ansätze für das Arbeiten mit Ermüdungsdaten basierend auf dem Ermüdungsbruch diskutiert. Dabei werden mehrere statistische Modelle eingeführt, analysiert und an einen Datensatz aus der Literatur angepasst.

Unter diesen Modellen finden sich zwei oft verwendete Modelle wieder, das Lineare Modell und das Generalisierte Lineare Modell, welche in der Praxis oft noch den Status Quo darstellen. Darüber hinaus werden wir zwei alternative Ansätze besprechen. Beim ersten Modell modellieren wir die Parameter der gewählten Verteilung basierend auf Charakteristika von Ermüdungsdaten, es wird Parameter Regression Model genannt. Der zweite Ansatz zielt darauf aus, das Problem weniger Datenmessungen zu eliminieren oder zumindest abzuschwächen, indem die Beobachtungen zu einer einheitlichen Verteilung hin transformiert werden. Das entsprechende Modell wird Standardizing Approach Model genannt.

Ein Vergleich der eingeführten Modelle mit der freien Statistiksoftware R basierend auf deren Präzision beim darstellen des Medians und der Streuung des Datensatzes führt uns zu einem Ranking. Bei diesem wird das Standardizing Approach Model vor dem Generalisierten Linearen Modell als erstes gereiht. Den dritten Platz teilen sich das Parameter Regression Model und das Lineare Modell mit drei Prädiktoren. Das letztgereihte Modell ist das Lineare Modell mit zwei Prädiktoren. Sämtliche statistische Analysen wurden mit dem Programmpaket R (Version 3.2.3) durchgeführt.

ACKNOWLEDGEMENT

I would first like to thank my thesis advisor Univ.-Prof.i.R. Dipl.-Ing. Dr.techn. Ernst Stadlober of the Institute of Statistic at Graz University of Technology and Univ.-Prof. Mag. Dr.rer.nat. Erika Hausenblas of the Department Mathematics and Information Technology at Montanuniversitaet Leoben. There was always a helping hand and a friendly ear whenever I needed one.

I would also like to thank Dr. Hans-Peter Gänser of the Materials Center Leoben who contributed his mechanical engineering expert knowledge and steered me into the right direction. Without his help and input, this thesis would not be what it is now.

Finally, I must express my very profound gratitude to my parents and to my brothers and my sister for providing me with unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them. Thank you.

Matthias Thomas Maurer

Contents

List of Figures

List of Tables

1. Introduction

This thesis discusses different statistical approaches to fatigue data focused on the total life approach. Different statistical models are introduced, analysed, and used to model an exemplary data set from the literature.

This first chapter gives an insight on the thesis structure. The chapters are described shortly and the exemplary data set is introduced. Findings of the author using this data set are also described.

In chapter two we introduce the fatigue phenomenon, review the history of fatigue research, and identify different approaches to fatigue. These insights explain, why we are interested in this field of research. Starting with safety concerns of mining and rail-road companies different theories of fatigue were discussed. Also the tools for analysing this problem, such as the Wöhler or S-N curve are introduced in the process. From a presented separation of fatigue approaches, we then focus on the total life approach. Furthermore, we explain how exactly stress is applied in the Wöhler experiment, one case of application in the context of the total life approach.

The third chapter introduces regression analysis and two very common regression analysis model types, the Linear Model and the Generalized Linear Model. For both of these models, we discuss the model definitions and assumptions. Furthermore, we define the model parameter estimates and introduce hypothesis tests in this context. They are still widely used models due to their easy handling and simple interpretation and hence represent the status quo.

Chapter four introduces statistical fundamentals, such as estimators, confidence intervals, and hypothesis tests. The formal definitions are used to build a common ground for further discussions. We used examples to illustrate the definition and deduce estimators, confidence intervals, and hypothesis tests for our practical applications.

In chapter five we elaborate a methodology for analysing fatigue data in a defined form with parameter regression. To visualise the procedure we use an exemplary data set from the literature. Starting with moment modelling we proceed to parameter modelling for particular distributions. This modelling approach is then analysed by using simulations for different distributions in combination with multiple fitting methods.

The sixth chapter introduces a methodology for standardizing the given data points from different stress levels of the Wöhler curve and working with the much bigger, standardized dataset. This allows us to calculate estimates and quantiles more efficiently. This approach is intended to reduce the disadvantage of the low number of data points in the context of fatigue data due to high experiment costs.

Chapter seven is used to apply the introduced models from chapter two to our exemplary data set from literature. Thereby, we used axis transformations to provide conditions suitable for the model assumptions. This leads us to different models providing confidence intervals and prediction intervals for the given fatigue data.

In chapter eight we compare the different models introduced in the previous chapters. This includes two Linear Models, one Generalized Linear Model, a Parameter Regression Model adjusted to fatigue data, and a Standardizing Approach Model. Two aspects are used for the comparison - the ability to fit the median and the ability to describe the scatter.

1.1 The Dataset

In this thesis we are working with the AAW dataset which is taken from Shen (1994, pp. 259 ff.) and can be found in Table 1. It includes two hundred observations, with twenty observations for each of the ten stress levels.

Table 1: Annealed Aluminium Wire (AAW) fatigue data

stress level[MPa]									
294.3	220.7	176.6	134.9	105.4	83.4	73.6	56.4	54	51.5
100 cycles to failure									
53	51	62	91	128	182	120	1140	2850	8200
62	61	94	93	156	250	400	1300	3080	8390
65	70	100	94	174	257	450	1570	3360	9380
66	77	100	97	190	286	480	1570	3770	10240
70	86	102	145	190	290	620	1590	3800	10400
75	90	108	159	197	337	650	1700	3960	10480
80	91	113	160	200	350	650	1800	4270	11000
84	93	126	162	210	364	670	2010	4970	11030
87	96	128	179	213	399	700	2050	5100	11360
88	97	139	185	244	400	800	2100	5510	11450
90	97	140	198	251	407	810	2300	5600	11470
92	101	142	208	254	440	830	2440	5950	11500
92	103	143	210	267	451	880	2500	6170	11510
94	112	147	218	268	460	910	2510	6600	11630
95	115	151	221	269	461	920	2570	6680	12000
95	116	152	224	283	468	940	2660	6850	12100
98	123	166	224	285	487	950	2730	7140	13190
100	125	169	257	295	500	1040	2870	7330	13200
105	134	170	258	309	543	1080	2960	8490	13210
118	159	182	278	382	556	1120	3090	8950	16300

This data is taken from Shen (1994, pp. 262), Table $D.1$ and describes a fatigue test conducted with ten stress-levels, each with 20 iterations. It is originally released by Freudenthal (1952).

Shen (1994) used this dataset among others to develop methods for providing a statistical summary of material fatigue stress-life data for engineering design purposes within his PhD thesis. As significant achievements, he described the following statements:

- The bilinear model seems to provide a consistently adequate description of the trend of fatigue data, using representative fatigue data sets.
- The pure X error sources model seems to provide a consistently adequate description of the uncertainties observed in heteroscedastic fatigue data (like the AAW dataset represented in Table 1). The pure X error source model is based on recognition of the uncertainties in local fatigue stress.

For understanding this achievements we first need to introduce the bilinear model and the pure X error source model. The bilinear model represent a model for the mean μ_Y and is defined by the implicit equation

$$x = a_1 + a_2(\mu_Y + a_5\sqrt{(\mu_Y - a_3)^2 + a_4}),$$

with the parameter vector $a = (a_1, a_2, a_3, a_4, a_5)$, the stress (or transformed stress) x and the mean μ_Y for the response Y (cycles to failure or transformed cycles to failure). For the pure X error source model we first have to have a look at the model defined by the equation

$$Y = y_L(x + \epsilon_{x_L}; a) + \epsilon_{y_L}. \tag{1.1}$$

This equation models the response Y (cycles to failure or transformed cycles to failure) as function y_L of the stress (or transformed stress) $x + \epsilon_{x_L}$ and the model parameter vector a added to the error term ϵ_{y_L}. The stress (or transformed stress) $x + \epsilon_{x_L}$ consists of the mean x and an error term ϵ_{x_L}. This X error source is explained by different disseminated stresses on the material sample, which lead to an unknown stress at the point of crack initialisation. A pure X error source model is now characterised by the equation

$$Y = y_L(x + \epsilon_{x_L}; a),$$

which result from Equation (1.1) when setting ϵ_{y_L} to zero. It is furthermore assumed that $\epsilon_{x_L} \sim N(0, \sigma^2_{x_L})$ holds.

2. Fatigue

2.1 The fatigue phenomenon

Fatigue is the phenomenon describing the damage and failure of materials under cyclic stress (Suresh, 1998). In our context, the main focus of fatigue is directed towards the failure of materials. The cyclic stress causes material failures with a much lower peak value than the safe load determined. The form of fatigue, as well as the materials which are effected by fatigue, are numerous.

There are multiple forms of fatigue failure, such as mechanical fatigue, creep-fatigue, thermomechanical fatigue, corrosion fatigue, sliding contact fatigue, rolling contact fatigue, and fretting fatigue (Suresh, 1998). Mechanical fatigue is based on mere fluctuations in externally applied stresses; Creep-fatigue is the combination of mechanical fatigue and high temperatures; Thermomechanical fatigue describes creep-fatigue with fluctuating temperature; Corrosion fatigue takes place in the presence of chemically aggressive environments; Sliding contact fatigue or rolling contact fatigue emerge in conjunction with stress applied to sliding or rolling contact between materials Fretting fatigue describes the failure due to pulsating stresses along with oscillating motion and frictional sliding between surfaces.

The number of materials affected by fatigue is considerable high (Suresh, 1998). Due to the nature of fatigue, it occurs everywhere cyclic stress is applied, no matter if the materials are metallic or non-metallic, brittle or ductile, monolithic or composite, natural or synthetic. The main focus of the fatigue research is directed towards metallic materials; although, there is an interest also in ceramic, polymers, and their composites. This trend starting in the 1990s is motivated in mechanical, thermal and environmental characteristics of these materials unobtainable in conventional metals.

2.2 History of fatigue research

Fatigue research dates back to the first half of the nineteenth century based on published reports (Suresh, 1998). Starting with studies about metal fatigue in mining (Albert, 1838), an increasing interest in this phenomenon was observable animated by the emerging use of ferrous structures, especially in the railway industry. The Institution of Mechanical Engineers in Britain started to explain the fatigue phenomenon with the so-called crystallization theory. According to this theory, the crystallization of the underlying microstructure caused the fatigue failure. In this context, August Wöhler characterized fatigue behaviour in terms of the still used and very famous stress amplitude-life (S–N) curves.

Wöhler's S-N curve is showing the result of the so-called Wöhler experiment originally using a rotating bending machine (Suresh, 1998). This machine is used to expose a test sample to cycling rotation and/or bending stress with a fixed maximum peak value, called the stress level. The number of cycles until the failure occurs is denoted as cycles to failure. Due to time and cost constraints, the number of cycles is lim-

ited; A sample surviving a large, fixed number of cycles is called a run-out sample. This is done for multiple stress levels and multiple test samples per stress level. The resulting values are plotted as points - stress level (S) against the logarithmic scale of cycles to failure (N). A smooth curve representation of the resulting scatter plot is called Wöhler or S-N curve and can be seen in Figure 1. In this figure the endurance limit σ_e is shown; It characterizes the stress level below which a material is expected to have no fatigue failure.

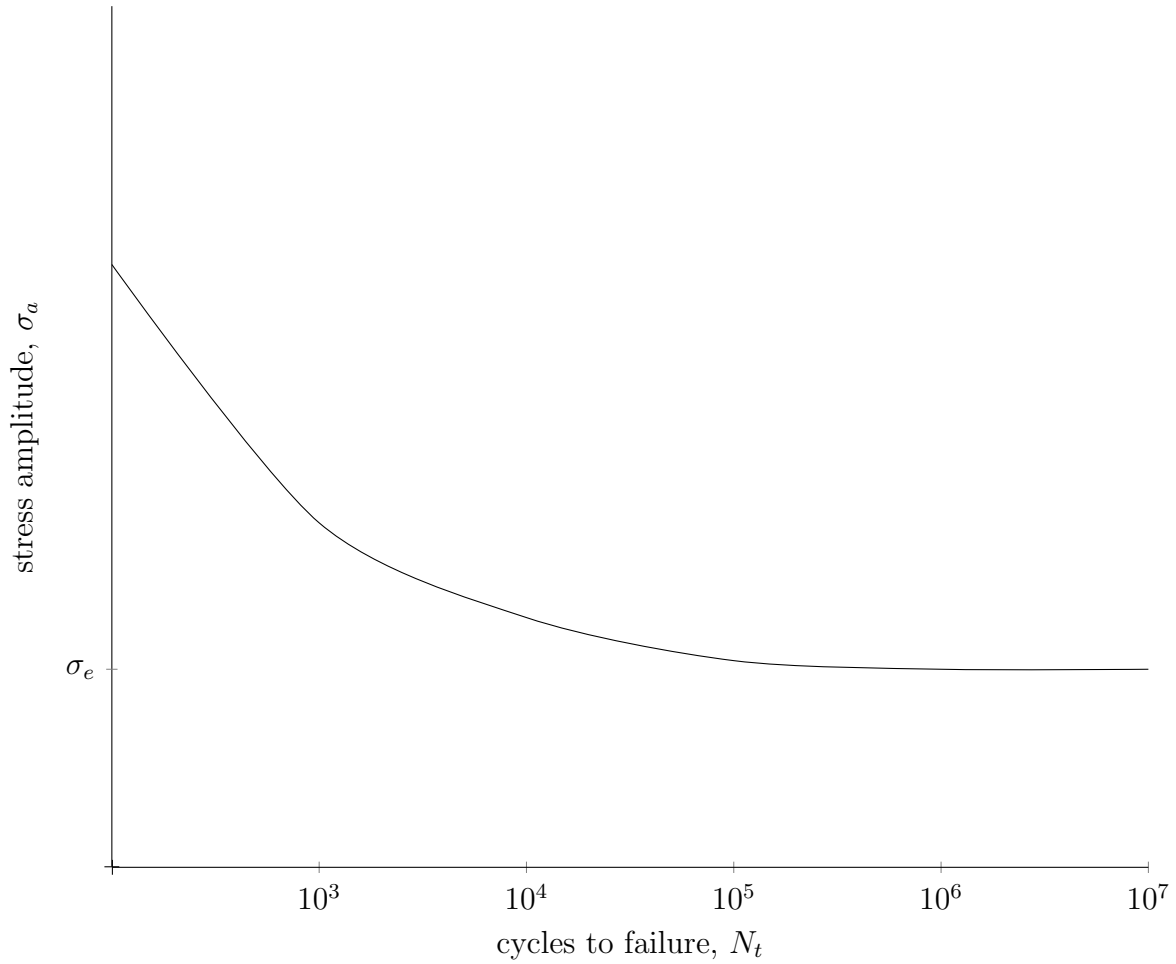


Figure 1: Typical S-N curve showing the total fatigue life of a nominally smooth-surfaced, 'defect-free' material (Suresh, 1998). The total fatigue life is defined as the sum of the number of cycles to initiate a fatigue crack and the number of cycles to propagate it subcritically to some final crack size.

The crystallization theory about the fatigue phenomenon was disproved and hence laid to rest based on optical micrographs of cyclic damage on the sample surface (Suresh, 1998), starting in 1900. The examined samples showed cracks with the progression of the fatigue deformation (Ewing and Rosenhain, 1899). A single major

crack leads to the fatigue failure in the sequel.

By the 1930s multiple properties of the fatigue behaviour were examined; the fatigue phenomenon becomes a major scientific research field (Suresh, 1998). Basquin, for example, showed, that a log-log plot of the stress level against the number of circles to failure reflects the fatigue behaviour of metals in a wide stress range (Basquin, 1910). Also the influence of vibrations and heat to the lifespan of materials were discussed.

The linear elastic fracture mechanics approach was initiated by Irwin (1997) in 1957, who take up the crack growth observation during the fatigue deformation from the early 1900s (Suresh, 1998). He showed, that the force forming a crack could be expressed in terms of the scalar quantity known as the stress intensity factor (K). This lead to the assumption, that the increment of fatigue crack advance per stress cycle (da/dN) could be related to the range of the stress intensity factor (ΔK) (Paris et al., 1961). A major advantage of this technique is, that no detailed information about the fatigue crack mechanisms is needed.

Researchers directed their attention increasingly on the mechanisms of subcritical crack growth due to the introduction of the electron microscopy (Suresh, 1998), starting in the 1960s. This microscopy was able to shine a light on various mechanical, microstructural and environmental factors related to fatigue deformation, crack initiation, and crack growth. Among others, Suresh and Moss (1980) categorized the basic crack properties and deduced further information for them.

The geometrical conditions of fatigue cracks moved in the research focus (Suresh, 1998). Pearson (1975) identified the 'short crack problem' in 1975, describing, that comparable small cracks show a higher crack growth than tall cracks when stressed with the same value of ΔK. It also occurs, that the crack growth rate diminishes with an increasing crack length. This phenomenon points towards a geometrical dependence of the crack growth.

Fatigue research creates models to estimate material's useful fatigue life, so called life prediction models (Suresh, 1998). Such models include at least fatigue failure under fixed amplitudes of cyclic stress. More complex models need to handle variable and multi-axial stress amplitudes, corrosive environments, and extreme temperatures since structural components are exposed to these service conditions. The model creation for such complex conditions is one of the hardest tasks in fatigue research.

2.3 Approaches to fatigue

There are different approaches to fatigue, they can be categorized into 'total life approaches', 'defect-tolerant approaches', 'safe-life' and 'fail-safe' approaches (Suresh, 1998). These approaches incorporate the role of crack initiation and crack growth differently in the calculation of useful fatigue life. Also, the useful fatigue live defini-

tion varies from one approach to the other.

Total life approaches account for the total life of material samples; It is defined as the number of cycles to crack initiation plus the number of cycles until failure occurs (Suresh, 1998). The total life cycles to failure are determined using initially uncracked samples under controlled stress amplitudes. In this context, the S-N curve is a famous example of data representation, shown in Figure 1.

The defect-tolerant approach not only takes the cycles to failure into account but rather the circles until a critical crack size emerges in the material (Suresh, 1998). The basic idea is, that all materials do have cracks which propagate under stress cycles. Fatigue life is defined as the number of cycles until a sample shows cracks of critical size. This conservative approach has been used in critical applications when a failure result in the loss of human lives.

Safe-life and fail-safe approaches were created by aerospace engineers for their special requirements to failure avoidance (Suresh, 1998). For the first one, the sample is tested in the laboratory using the typical conditions for its field of application. The fatigue life estimation, gained from this approach is then modified with a safety factor to obtain the component's 'safe-life' prediction. Components are exchanged after reaching the safe-life, independent of their condition. This approach depends on the prevention of an initial crack. The fail-safe concept, on the other hand, grounds in the idea, that the failure of one component should not destroy the structural integrity of the overall structure. Damaged components are replaced within a regular inspection of the structure.

2.4 Stress amplitude categorization

For a generic material sample, one has to differentiate three normal stresses and six shear stresses (Schlottmann and Schnegas, 2016). The stresses are represented as tensors and can be transformed into main stress tensors without shear stresses for homogeneous and isotropic materials, see Figure 2. A matrix representation of this situation is given in Equation (2.1).

$$\sigma^T = \begin{pmatrix} \sigma_x & \tau_{xy} & \tau_{xz} \\ \tau_{yx} & \sigma_y & \tau_{yz} \\ \tau_{zx} & \tau_{zy} & \sigma_z \end{pmatrix} \rightarrow \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \tag{2.1}$$

The most simple geometric form of a material sample is the bar (Schlottmann and Schnegas, 2016). In a bar-shaped sample there are the quantities axial force $F_a(s)$, bending moment $M_b(s)$, torsional moment $M_t(s)$, and shear force $F_s(s)$, see Figure 2. These quantities result in the stresses σ_a, σ_b, τ_t, and τ_s; The relation between forces/moments and stresses and also the related stress tensors are listed in Table 4.

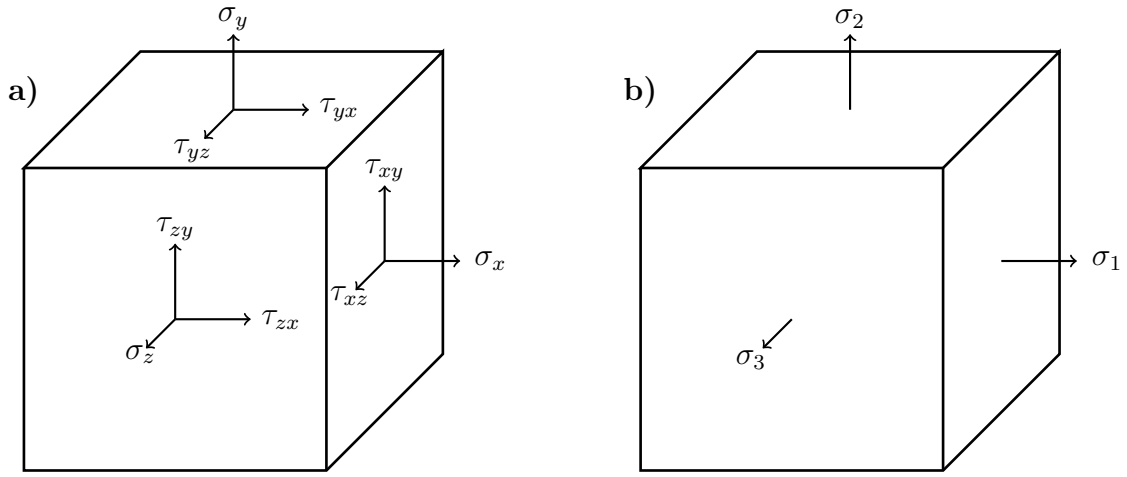


Figure 2: a) Normal and shear stresses b) Main stresses observable on a material sample (Schlottmann and Schnegas, 2016).

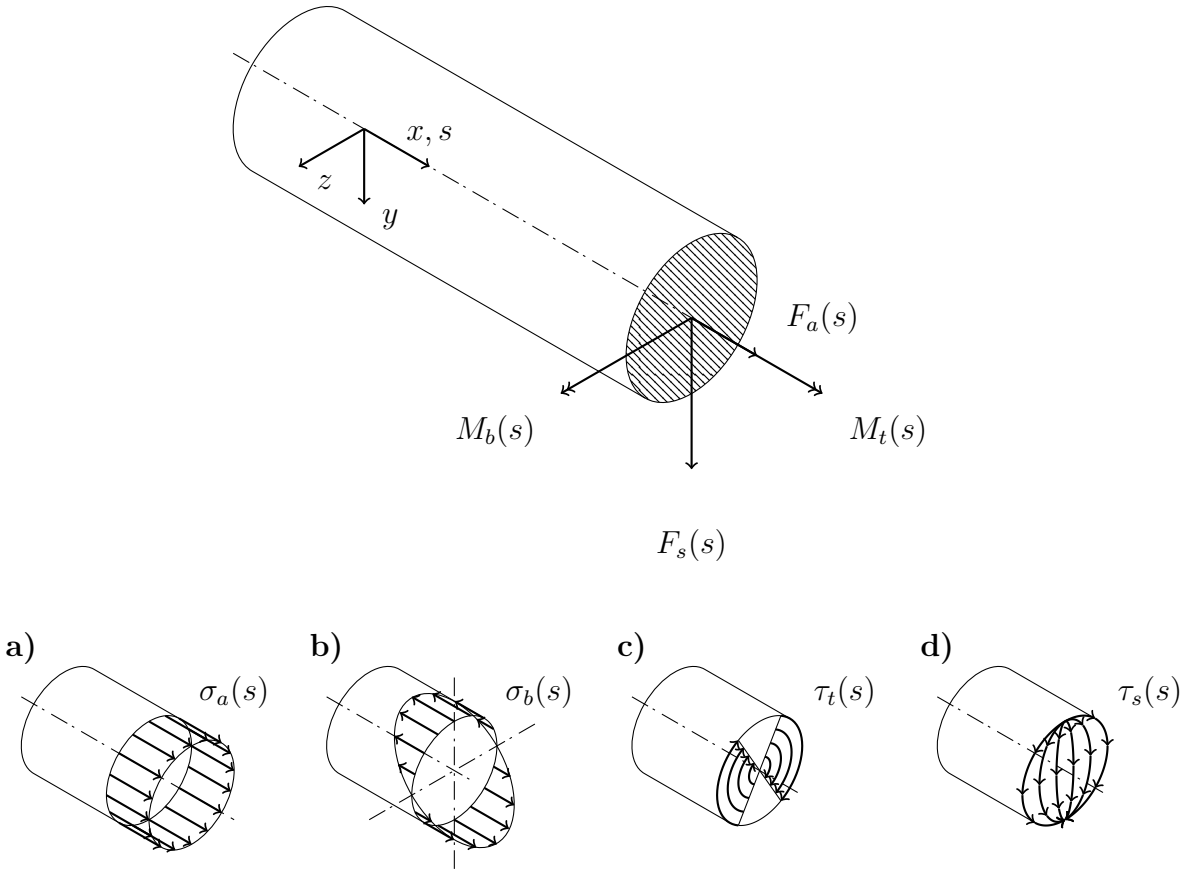
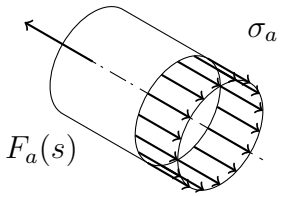
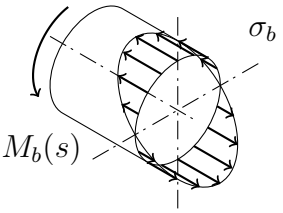
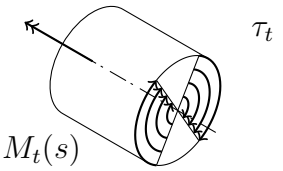
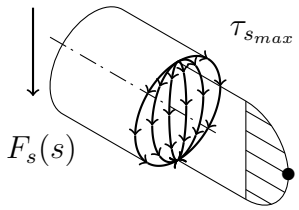


Figure 3: Stress in bar-shaped material samples (Schlottmann and Schnegas, 2016). a) Stress based on axial force $F_a(s)$; b) Stress based on bending moment $M_b(s)$; c) Stress based on torsional moment $M_t(s)$; d) Stress based on shear force $F_s(s)$

Load (force/moment)	Visualisation	Calculation	Stress tensor
Axial force $F_a(s)$		$\sigma_a = \frac{F_a(s)}{A}$	$\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $\sigma_1 = \sigma_a$
Bending moment $M_b(s)$		$\sigma_b = \frac{M_b(s)}{W_z}$ $\sigma_b = \frac{M_b(s)}{I_z} \cdot y$	$\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $\sigma_1 = \sigma_b$
Torsional moment $M_t(s)$		$\tau_t = \frac{M_t(s)}{W_t}$ $\tau_t = \frac{M_t(s)}{I_t} \cdot r$	$\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $\sigma_1 = \lvert\tau_t\rvert$ $\sigma_1 = -\sigma_2$
Shear force $F_s(s)$		$\tau_{s_{max}} = \chi \frac{F_s}{A}$	$\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $\sigma_1 = \lvert\tau_{s_{max}}\rvert$ $\sigma_1 = -\sigma_2$

A $\cdots$ cross sectional area
I_z, I_t $\cdots$ inertia torque
W_t $\cdots$ moment of resistance
χ $\cdots$ ratio of maximum shear stress to nominal shear stress

Figure 4: Basic stresses in bar-shaped material samples (Schlottmann and Schnegas, 2016).

Besides the bar-shaped form and the loads shown in Table 4, there is a list of geometric forms in combination with applied loads used as model cases (Schlottmann and Schnegas, 2016). These so-called stress states can be categorized into rotationally symmetric states, shell states, material states including a notch, and rolling/sliding contact states. Examples for the rotationally symmetric case are the cylinder, the hole in an endless plate, the thin pipe, or the spherical shell. Shell states are for example the plate or the shell under load. Exemplary forms of notches are cycles and parabolas. Rolling/Sliding contact states are, for example, two spheres, two cylinders, or the bolt (ingrained cylinder) under load.

An essential consideration when talking about fatigue is the stress behaviour over time (Schlottmann and Schnegas, 2016). Besides the cyclic stress, which constitutes the most important stress behaviour, also static stress and stochastic stress behaviour are considered. Cyclic stress is characterized by a periodical stress over time curve; The probably most famous example for a periodical curve is the sinus curve. Static stress is defined by a constant stress amplitude over time. Stochastic stress is present when the correlated stress over time curve is random within given limitations.

The stress range is divided into three areas for each material sample based on the Wöhler curve (Schlottmann and Schnegas, 2016). The value range up to $10^3...10^4$ stress cycles is called Low-Cycle-Fatigue (LCF) area. The value range starting with $10^6...10^7$ stress cycles is called Very-High-Cycle-Fatigue (VHCF) area. The value range in between is called High-Cycle-Fatigue (HCF) area.

3. Regression Analysis

3.1 What is regression analysis

Regression analysis is a technique to identify the main features of relationships between quantities in a system based on measured data (Draper and Smith, 2014). With the knowledge of how a variable influences another variable within the same system, we are either able to adjust the system to our needs or to make predictions among other things. In the first case, this might include the effect the working temperature has on the quality of a product in a production line; Then we can manipulate the temperature to gain the best possible product quality. In the last mentioned case, we could use air measurements to predict the rain possibility for the next day.

Two main types of variables are used in this setting; the predictor variables and the response variables (Draper and Smith, 2014). Predictor variables are those, who are observed or can be set to a value within the considered system. The response variables, in contrast, are the variables influenced by the change of the predictor variables. In the earlier production line example, the working temperature is the predictor variable and the product quality is the response variable.

The type of relationship ranges from a simple linear relation to a much more complex behaviour, which can not be reproduced easily (Draper and Smith, 2014). A linear relation between two variables in this context does not refer to linearity in the proper sense like used when talking about a linear function. This is due to the probabilistic nature of the observed variables; Instead of the relation $y = f(x)$ in case of two variables we consider the relation $y = f(x) + \epsilon$, with ϵ as a random error and the so-called model function $f(x)$ as e.g. linear function. This is also true in case x, y and ϵ are vectors. For complex relations, we can apply approximations using simple mathematical functions, e.g. polynomials.

The model function is generally of a predefined form and depends on the predictor variables and a parameter vector, the random error distribution type is known (Draper and Smith, 2014). The parameters are estimated from the data, meaning the observations for response and predictor variables. This process of determining the parameters is called fitting the model to given data. The random error is often assumed to be normally distributed with mean zero. It is also common to assume independent random errors.

3.2 The Linear Model and its extensions

A very popular series of statistical models are represented by the Linear Model and its extensions, the Generalized Linear Model, and the Generalized Additive Model (Wood, 2006). They vary in their ability to model data based on the basic model assumptions. The simpler models (such as the Linear Model) have very inflexible model assumptions but can offer a wide variety of theoretical derivations and well-

founded statements. With rising level of complexity, like in the Generalized Linear Model or even in the Generalized Additive Model we are able to model more and more relationships between data, but increasingly loose the potential for well-founded derivations. We will talk about the first two of these models.

A linear model (LM) is a statistical model for a univariate response variable, which is expressed as the linear combination of predictor variables (called linear predictor) and a random error with mean zero (Wood, 2006). Linearity in this context means, that the linear predictor depends linearly on the parameters, which are the coefficients of the linear combination. This model is widely used in most science branches for modelling tasks. It is an easy to use model, with elegant theory.

The Generalized Linear Model (GLM) extends the LM by relaxing the linearity and distribution assumption (Wood, 2006). Instead of assuming the expected value of the response to be the linear predictor, we are now assuming it as a smooth monotonic function of the linear predictor. The normal distribution assumption is replaced by the exponential family distribution assumption. The exponential family embeds multiple distributions, such as the Normal, Poisson, Binomial, and Gamma distribution.

3.2.1 Linear Model

The linear model is described with the following equation (Wood, 2006):

$$\mu = X\beta, \quad y \sim N(\mu, I_n\sigma^2), \quad y = \mu + \epsilon, \tag{3.1}$$

whereas μ represents the expectation vector of the response vector y, X stands for the model matrix including the predictor variables with n rows and p columns, β describes the p-dimensional model parameter vector, I_n constitutes the identity matrix of rank n, and σ represents the constant model variance parameter. Equation (3.1) embeds multiple observations; the i-th entry of the vector μ in combination with the i-th row of the model matrix and the parameter vector β stands for one observation. The statement can be rewritten as:

$$\mu_i = X_i\beta, \quad y_i \sim N(\mu_i, \sigma^2), \quad y_i \text{ independent}, \quad 1 \leq i \leq n,$$

whereas $X = (x_1 x_2 \cdots x_n)^T$ and row vectors $x_i = (x_{i,1}, \cdots, x_{i,p})$. The structure of the X_i row vectors define the model properties. For example, a constant value $x_{i,j} = 1 \quad 1 \leq i \leq n, j = j_0$ adds a constant summand into the expectation term. A classification into p different categories, or cells can be gained by including p columns into the model matrix with $x_{i,j} = 1$, if the i-th observation is in the j-th cell or class, and $x_{i,j} = 0$ otherwise. Each column in the model matrix represents the realisation of one predictor, each row in the model matrix represents one statistical entity. The vector y represents the realisations of the response variable Y.

The least square method is used to obtain an estimate for the parameter vector β (Wood, 2006). Starting with n observations $(y_i, x_{i,1}, \cdots, x_{i,p})$ for $1 \leq i \leq n$ one can

calculate the squared error term $\mathcal{S}$ in dependency of β:

$$\mathcal{S}(\beta) = \sum_{i=1}^{n} (y_i - \mu_i(\beta))^2 = \|y - X\beta\|^2, \tag{3.2}$$

with $y = (y_1, \cdots, y_n)$. Minimizing the term (3.2) with respect to β leads to the point estimator $\hat{\beta}$, which can be expressed as (see Theorem A.31):

$$\hat{\beta} = \arg\min_{\beta} \mathcal{S}(\beta) = (X^T X)^{-1} X^T y. \tag{3.3}$$

The least square parameter estimate $\hat{\beta}$ is normally distributed (Wood, 2006). It's expected value equals the true parameter vector β, hence it is unbiased. Since $\hat{\beta}$ is a linear combination of normal random variables, it is a normal random variable itself. Additionally its covariance matrix equals $(X^T X)^{-1}\sigma^2$. To sum up β's distribution properties one can write (see Theorem A.32)

$$\hat{\beta} \sim N(\beta, (X^T X)^{-1}\sigma^2).$$

An estimation for the model parameter σ can be calculated as (Wood, 2006):

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \left(y_i - \mu_i(\hat{\beta}) \right)^2 = \frac{1}{n-p} \|y - X\hat{\beta}\|^2 = \frac{\mathcal{S}(\hat{\beta})}{n-p}.$$

This estimator is independent of the estimator $\hat{\beta}$ (due to the independence of $\hat{\beta}$ and $\mathcal{S}(\hat{\beta})$, see Theorem A.33) and can be used to build a hypothesis test for single parameter estimates $\hat{\beta}_i$ ($H_0 : \beta_i = 0$). This is possible, since $(n-p)\frac{\hat{\sigma}^2}{\sigma^2}$ is chi-squared distributed with $(n-p)$ degrees of freedom (see Theorem A.34). Consequently, $\frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_i}}}$ is t-distributed with $(n-p)$ degrees of freedom, whereas $\hat{\sigma}^2_{\hat{\beta}_i}$ describes the i-th diagonal element of $\hat{\beta}$'s covariance matrix $(X^T X)^{-1}\sigma^2$ (see Proposition A.35).

A central role is assigned to the so called influence or hat matrix H, which transforms a data vector y into fitted values $\hat{\mu} = Hy$ (Wood, 2006). Due to the Equation (3.1) and Equation (3.3) we can write:

$$\hat{\mu} = X\hat{\beta} = X(X^T X)^{-1} X^T y = Hy, \quad H := X(X^T X)^{-1} X^T.$$

The diagonal element h_{ii} of the hat matrix H is also called the leverage of the i-th observation.

A useful quantity to check the model base assumption is the residual (Draper and Smith, 2014). It is defined as the difference between the observed value and the fitted value, $r_i = y_i - \hat{\mu}_i$. The vector of residuals r can be written as $r = y - \hat{\mu} = y - Hy = (I - H)y$, hence it is a normal distributed random vector (as combination of normal distributed random vectors, see Example A.28). It holds $r \sim N(0, \sigma^2(I - H))$. By

standardizing we gain the standardized residual $d = \frac{r}{\hat{\sigma}}$. The i-th studentized residual r_i^* for $1 \leq i \leq n$ is given as:

$$r_i^* = \frac{r_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}},$$

with expectation zero and variance one.

The quantities R^2 and R^2_{adj} can be used to compare model adequacy (Draper and Smith, 2014). They are defined by:

$$R^2 = 1 - \frac{\sum_{i=1}^n r_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{with} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

and

$$R^2_{adj} = 1 - \frac{n-1}{n-p} \Big(1 - R^2 \Big).$$

The denominator depends only on the observations and is independent of the chosen model. The numerator, on the other hand, depends on the model and will not rise when adding another predictor. Hence R^2 will decrease or stay the same when adding another predictor to the model. R^2_{adj} also considers the number of used predictors in the model and might also decrease when adding another predictor. One could start with the model only containing the intercept and adding predictors till R^2_{adj} is maximal or till the change is very small. This approach is called forward selection. When starting with a model containing the largest number of predictors and removing them till R^2_{adj} reaches the maximum or does not change too much, a backward selection is done. Comparing those quantities seems reasonable within the same set of predictors and responses.

The influence of an observation on the parameter estimate $\hat{\beta}$ can be measured using the Cook's Distance D_i (Cook, 1977; Draper and Smith, 2014). It is defined for the i-th observation as:

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})^T X^T X (\hat{\beta}_{(i)} - \hat{\beta})}{p\hat{\sigma}^2} = \frac{(r_i^*)^2 h_{ii}}{p(1 - h_{ii})},$$

whereas $\hat{\beta}_{(i)}$ describes the parameter estimator for β in the model ignoring observation i and h_{ii} describes the i-th diagonal element of the hat matrix H. Montgomery et al. (2012) recommend to classify values larger than one as influential.

The F-ratio test can be used to test if several model parameters are simultaneous zero (Wood, 2006). Therefore we have a look at two estimates. First, we have the *full* model estimate $\hat{\beta} \in \mathbb{R}^p$ including all p parameters. Second, we have the *reduced* model estimate $\dot{\beta} \in \mathbb{R}^{p-q}$ including all parameters from the full model, but the q tested ones. Then we can write for the test statistic (see Theorem A.36):

$$
\begin{aligned}
F(\hat{\beta}, \dot{\beta}) =& \frac{n-p}{q} \cdot \frac{\sum_{i=1}^{n}(y_i - \dot{\mu}_i)^2 - \sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2}{\sum_{i=1}^{n}(y_i - \hat{\mu}_i)^2} \\
=& \frac{n-p}{q} \cdot \frac{\|y - X\dot{\beta}\|^2 - \|y - X\hat{\beta}\|^2}{\|y - X\hat{\beta}\|^2} \\
=& \frac{n-p}{q} \cdot \frac{\mathcal{S}(\dot{\beta}) - \mathcal{S}(\hat{\beta})}{\mathcal{S}(\hat{\beta})} \sim F_{q,n-p}.
\end{aligned} \tag{3.4}
$$

If the value $F(\hat{\beta}, \dot{\beta})$ is now larger then the reference value $f_{q,n-p,1-\alpha}$ ($1-\alpha$ quantile of the $F_{q,n-p}$ distribution) we need to work with the full model estimate $\hat{\beta}$, otherwise the reduced model estimate $\dot{\beta}$ is sufficient. In this setting the null hypothesis is, that taking the reduced model is sufficient. The value of α describes the accepted type 1 error probability; This is the probability, that the null hypothesis is wrongly rejected.

3.2.2 Generalized Linear Model

A GLM is given with (Wood, 2006):

$$g(\mu_i) = X_i\beta, \quad Y_i \sim Exp(\theta_i, \phi)$$

with $\mu_i = \mathbb{E}[Y_i]$, g as smooth monotonic 'link function', $X = (x_1 x_2 \cdots x_n)^T$ as model matrix with rows $x_i = (1, x_{i,1}, \cdots, x_{i,p})$ $(1 \leq i \leq n)$, β as parameter vector, and $Exp(\theta_i, \phi)$ as distribution from the exponential family.

The exponential family used in GLMs describes a wide range of distributions (Wood, 2006). A distribution is member of the exponential family, if it's density or probability function f can be written as:

$$f(y, \theta) = e^{\frac{y\theta - b(\theta)}{a(\phi)} + h(y,\phi)},$$

whereas $a(\cdot), b(\cdot)$ and $h(\cdot,\cdot)$ describe real, known functions. The scale parameter ϕ is a known value for which $a(\phi) > 0$ holds, and θ is a known parameter. We write $f \in Exp(\theta, \phi)$ for the distribution f as member of the exponential family. Examples for member of the exponential family are Normal-, Binomial-, Poisson-, Exponential-, and Gamma-distribution.

Example 3.1. ***(Poisson distribution)***
The discrete Poisson distribution, defined for random variable $X \sim Poi(\lambda)$ through $\mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$ for $k \in \mathbb{N}_{\geq 0}$ and $\lambda > 0$ is member of the exponential family. The corresponding probability function can be rewritten as:

$$
\begin{aligned}
\mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} =& e^{k\cdot\log(\lambda) - \lambda - \log(k!)} \\
=& e^{\frac{k\cdot\theta - e^{\theta}}{1} - \log(k!)},
\end{aligned}
$$

for $a(\phi) = 1, \theta = \log(\lambda), b(\theta) = e^{\theta}$ and $h(k, \phi) = -\log(k!)$.

Example 3.2. ***(Gamma distribution)***
The Gamma distribution (see Definition A.16) is member of the exponential family. The density function can be rewritten as:

$$\begin{aligned} f(x) =& \frac{x^{k-1}e^{-x\lambda}\lambda^k}{\Gamma(k)} \\ =& e^{\frac{x(-\lambda/k)+\log(\lambda/k)}{1/k}+k\cdot\log(k)+(k-1)\log(x)-\log(\Gamma(k))} \\ =& e^{\frac{x\cdot\theta+\log(-\theta)}{\phi}+\frac{1}{\phi}\log\left(\frac{1}{\phi}\right)+\left(\frac{1}{\phi}-1\right)\log(x)-\log\left(\Gamma\left(\frac{1}{\phi}\right)\right)}, \end{aligned}$$

for $a(\phi) = \phi = \frac{1}{k}$, $\theta = -\frac{\lambda}{k}$, $b(\theta) = -\log(-\theta)$ *and* $h(x,\phi) = \frac{1}{\phi}\log(\frac{1}{\phi}) + (\frac{1}{\phi} - 1)\log(x) - \log(\Gamma(\frac{1}{\phi}))$.

For a exponential family member we can express certain distribution characteristics in form of distribution function elements (Wood, 2006). The expectation value μ can be expressed as $\mu = \mathbb{E}(y) = b'(\theta)$, and the variance $Var(y)$ can be written as $Var(y) = a(\phi)b''(\theta)$. Whereas the variance is rewritten as $Var(y) = a(\phi)b''(\theta) = a(\phi)V(\mu)$ with $V(\mu) = b''((b'(\mu))^{-1})$ as the so called variance function $V(\cdot)$ only depends on μ due to the identity $\theta = (b'(\mu))^{-1}$.

To obtain the parameter vector estimate $\hat{\beta}$ the Iterative Re-weighted Least Square (IRLS) method is used (Wood, 2006). Starting with β's log-likelihood function $l(\beta) = \sum_{i=1}^{n}\log[f_{\theta_i}(y_i)]$, we end up solving the equation

$$\sum_{i=1}^{n}\frac{(y_i-\mu_i)}{V(\mu_i)}\cdot\frac{\partial\mu_i}{\partial\beta_j} = 0 \quad \forall j.$$

This equation is solved by the following procedure (Nelder and Baker, 1972):

1) Set start values $\mu^{[0]}$ and $\eta^{[0]}$ for the iteration. Usual start values are $\mu_i^{[0]} = y_i$ and $\eta_i^{[0]} = g(\mu_i^{[0]})$.
2) Calculate the vector entries $z^{[k]}$ and the diagonal matrix entries $W^{[k]}$ as

$$\begin{aligned} z_i^{[k]} &= g'(\mu^{[k]})(y_i - \mu_i^{[k]}) + \eta_i^{[k]}, \\ W_{ii}^{[k]} &= \frac{1}{V(\mu_i^{[k]})g'(\mu_i^{[k]})^2}. \end{aligned}$$

3) Minimize the expression $\|\sqrt{W^{[k]}}(z^{[k]} - X\beta)\|^2$ with respect to β, to gain $\hat{\beta}^{[k+1]}$:

$$\hat{\beta}^{[k+1]} = \arg\min_{\beta}\|\sqrt{W^{[k]}}(z^{[k]} - X\beta)\|^2$$

These result is used to calculate the two vectors $\mu^{[k+1]}$ and $\eta^{[k+1]}$:

$$\begin{aligned} \eta^{[k+1]} &= X\hat{\beta}^{[k+1]}, \\ \mu_i^{[k+1]} &= g^{-1}(\eta_i^{[k+1]}) \quad 1 \le i \le n. \end{aligned}$$

An estimator for the scale parameter ϕ can be calculated based on the Pearson statistic in case it is unknown (Wood, 2006). Starting from the Pearson statistic, given by

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{V(\mu_i)},$$

we define the estimator $\hat{\phi}$ as:

$$\hat{\phi} = \hat{X}^2/(n-p). \tag{3.5}$$

The asymptotic distribution of the estimate gained from the IRLS procedure for known parameter ϕ can be stated (Wood, 2006). The estimator is asymptotically normal distributed with the following mean and variance:

$$\mathbb{E}(\hat{\beta}) = \beta$$

and

$$Var(\hat{\beta}) = (X^T W X)^{-1}\phi,$$

whereas $W = diag(w_i)$ is a diagonal matrix with $w_i = \left(V(\mu_i)g'(\mu_i)^2\right)^{-1}, 0 \leq i \leq n$. Consequently we get an asymptotic confidence interval $I(x)$ with confidence level $1-\alpha \in [0,1]$ for the expectation of the response:

$$I(x) = \left[g^{-1}\left(x^T\hat{\beta} - u_{1-\alpha/2}x^T\hat{Var}(\hat{\beta})x\right), g^{-1}\left(x^T\hat{\beta} + u_{1-\alpha/2}x^T\hat{Var}(\hat{\beta})x\right)\right],$$

when u_α is denoting the α-quantile of the standard normal distribution and $\hat{Var}(\hat{\beta})$ is denoting the evaluation of $Var(\hat{\beta})$ with $\hat{\mu}_i$.

A confidence interval for the expectation of the response can approximatively be given for unknown parameter ϕ (Wood, 2006) . Based on the assumption that $Y := \frac{\hat{\beta}-\beta}{\sqrt{\phi}} \sim N(0, (X^T W X)^{-1})$ and $U := \frac{\hat{\phi}(n-p)}{\phi} \sim \chi^2_{n-p}$ are independent, we gain the t-distributed random variable vector $T := \frac{Y}{\sqrt{U/(n-p)}} = \frac{\hat{\beta}-\beta}{\sqrt{\hat{\phi}}}$ with mean 0, covariance matrix $(X^T W X)^{-1}$, and $(n-p)$ degrees of freedom. This leads to the approximative confidence interval $I(x)$ with confidence level $\alpha \in [0,1]$:

$$I(x) = \left[g^{-1}\left(x^T\hat{\beta} - t_{n-p,1-\alpha/2}x^T(X^T\hat{W}X)^{-1}\hat{\phi}x\right), g^{-1}\left(x^T\hat{\beta} + t_{n-p,1-\alpha/2}x^T(X^T\hat{W}X)^{-1}\hat{\phi}x\right)\right],$$

when $t_{\nu,\alpha}$ is denoting the α-quantile of the t-distribution with ν degrees of freedom, $\hat{W}$ is denoting the evaluation of W with $\hat{\mu}_i$, and ϕ is the estimated scale parameter ϕ.

Model selection can be done using the so called Akaike Information Criterion (AIC) (Wood, 2006). This value is defined as $AIC(\hat{\beta}) = 2[-l(\hat{\beta}) + p]$ for known dispersion parameter ϕ and $AIC(\hat{\beta}) = 2[-l(\hat{\beta}) + p + 1]$ for unknown dispersion parameter ϕ; In this equation $l(\hat{\beta})$ stands for the maximized likelihood of the model corresponding to the parameter estimate $\hat{\beta}$ and p denotes the number of parameters in the model.

A generalized likelihood ratio test can be used to test if several model parameters are simultaneous zero in case the scale parameter ϕ is known (Wood, 2006). Examples for exponential family members with known scale parameter are the Poisson distribution and the Binomial distribution. Therefore we have a look at two estimates. First, we have the *full* model estimate $\hat{\beta} \in \mathbb{R}^p$ including all p parameters. Second, we have the *reduced* model estimate $\dot{\beta} \in \mathbb{R}^{p-q}$ including all parameters from the full model, but the q tested ones. If the null hypothesis is true, i.e. the reduced model is sufficient, then we have with $l(\tilde{\beta})$ for the maximized likelihood of the model corresponding to the parameter estimate $\tilde{\beta}$:

$$C(\hat{\beta}, \dot{\beta}) = 2[l(\hat{\beta}) - l(\dot{\beta})] \sim \chi^2_q.$$

This means, that we have to take the full model, if $C(\hat{\beta}, \dot{\beta})$ is higher then the reference value $c_{q,1-\alpha}$ ($1-\alpha$ quantile of the χ^2_q distribution). The value of α describes the accepted type 1 error probability; This is the probability, that the null hypothesis is wrongly rejected.

A hypothesis test based on the so called deviance can be used to test if several model parameters are simultaneous zero in case the scale parameter ϕ is unknown (Wood, 2006). The deviance $D(\hat{\beta})$ of a model with estimated parameter vector $\hat{\beta}$ is defined as:

$$\begin{aligned} D(\hat{\beta}) &= 2[l(\hat{\beta}_{max}) - l(\hat{\beta})]\phi \\ &= \sum_{i=1}^{n} 2w_i \Big[y_i(\hat{\theta}_i(\hat{\beta}_{max}) - \hat{\theta}_i(\hat{\beta})) - b(\hat{\theta}_i(\hat{\beta}_{max})) + b(\hat{\theta}_i(\hat{\beta})) \Big], \end{aligned}$$

with $w_i = \phi / a_i(\phi)$ and $\hat{\beta}_{max}$ as the parameter estimate of the saturated model: the model with as many parameters as data points ($n = p$). $\hat{\theta}^{max}$ and $\hat{\theta}$ describe the maximum likelihood estimates of the parameters for the saturated model and the model of interest ($\theta_i(\beta) = (b'(\mu_i(\beta)))^{-1} = (b'(g^{-1}(X_i\beta)))^{-1}$). If the null hypothesis is true, i.e. the reduced model is sufficient, then we have for $\hat{\beta}$ as the estimate for the full model (p parameters) and $\dot{\beta}$ the estimate for the reduced model ($p-q$ parameter) the approximate result:

$$F(\hat{\beta}, \dot{\beta}) = \frac{n-p}{q} \frac{D(\dot{\beta}) - D(\hat{\beta})}{D(\hat{\beta})} \dot{\sim} F_{q,n-p}.$$

If the value $F(\hat{\beta}, \dot{\beta})$ is now larger then the reference value $f_{q,n-p,1-\alpha}$ ($1-\alpha$ quantile of the $F_{q,n-p}$ distribution) we need to work with the full model estimate $\hat{\beta}$, otherwise the reduced model estimate $\dot{\beta}$ is sufficient. The value of $1-\alpha$ describes the accepted type 1 error probability; This is the probability, that the null hypothesis is wrongly rejected.

4. Statistical Fundamentals

The following chapter introduces concepts needed for the discussion of statistical models and their treatment, following Rüschendorf (2014). Starting with the definition of a statistical model we will discuss estimators and statistical tests. Those two concepts are central for working with statistical models, especially when it comes to predictions.

A statistical model or experiment $\mathcal{E} = (\mathcal{X}, \mathcal{A}, \mathcal{P})$ is formally defined as a triple, consisting of a base set $\mathcal{X}$, a sigma algebra $\mathcal{A}$ defined on this base set $\mathcal{X}$, and a parametric family of probability distributions $\mathcal{P}$. The pair $(\mathcal{X}, \mathcal{A})$ is forming a measurable space, an element of $\mathcal{X}$ is called sample. In contrast to a probability space, $\mathcal{P}$ does not only determine one probability measure but a range of possible probability measures.

As an example for a statistical model, we can name the Linear Model, defined in section 3.2.1. The base set is the set of the n-dimensional real numbers $\mathbb{R}^n$. The sigma algebra on $\mathbb{R}^n$ is defined by the Borel sigma algebra $\mathcal{B}(\mathbb{R}^n)$, the smallest sigma algebra containing all open subsets of $\mathbb{R}^n$. The parametric family of probability distributions $\mathcal{P}$ is given with the normal distribution family $\mathcal{P} = \{N(\mu, \sigma) | \mu \in \mathbb{R}, \sigma \in \mathbb{R}_{>0}\}$. The response vector y (as realization and not as random variable) is a sample.

4.1 Estimation Theory

A point estimator can be defined within a statistical model (Rüschendorf, 2014). Therefore we express the statistical model's family of probability distributions as $\mathcal{P} = \{P_\theta | \theta \in \Theta \subseteq \mathbb{R}^d\}$. In this setting any function $T : \mathcal{X} \to f(\Theta)$ is a point estimator for $f(\theta)$. A point estimator can be used to get insight about the unknown, true parameter vector θ, or functions of it. It is central, that a point estimator is in all relevant cases a random variable or random vector.

A desired feature of a point estimator is, that the expectation of its value equals the true parameter, i.e. it is unbiased. Therefore, we introduce the term bias of an estimator T estimating $f(\theta)$ as $Bias_\theta(T) = \mathbb{E}_\theta(T) - f(\theta)$. The estimator is called unbiased, if its bias is zero for all possible values of θ, i.e. $Bias_\theta(T) = 0 \quad \forall \theta \in \Theta$.

Example 4.1. ***(point estimator - normal distribution)***
Let $\mathcal{E} = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P})$ with $\mathcal{P} = \{P_\theta | \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_{>0}\}$. $\boldsymbol{X}_n = (X_1, ..., X_n)$ with $X_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ for unknown $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$ describes a sample from $\mathcal{E}$. We define the following two point estimator:

- $\hat{\mu} = T_\mu(\boldsymbol{X}_n) = \frac{1}{n} \sum_{i=1}^n X_i$,
- $\hat{\sigma}^2 = T_{\sigma^2}(\boldsymbol{X}_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - T_\mu(\boldsymbol{X}_n))^2 \quad or \quad \hat{\sigma} = T_\sigma(\boldsymbol{X}_n) = \sqrt{T_{\sigma^2}(\boldsymbol{X}_n)}$,

or equivalently $T = (T_\mu(\boldsymbol{X}_n), T_{\sigma^2}(\boldsymbol{X}_n)) : \mathbb{R}^n \to \Theta$. When we are looking at the expectation of these point estimators we can see, that they are unbiased:

$$\mathbb{E}(T_\mu(\boldsymbol{X}_n)) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} n\mu = \mu,$$

$$
\begin{aligned}
\mathbb{E}(T_{\sigma^2}(\boldsymbol{X}_n)) =& \frac{1}{n-1}\mathbb{E}\Big(\sum_{i=1}^{n}(X_i - T_\mu(\boldsymbol{X}_n))^2\Big) \\
=& \frac{1}{n-1}\mathbb{E}\Big(\sum_{i=1}^{n}(X_i^2 + T_\mu(\boldsymbol{X}_n)^2 - 2X_i^2 T_\mu(\boldsymbol{X}_n))\Big) \\
=& \frac{1}{n-1}\mathbb{E}\Big(\sum_{i=1}^{n}X_i^2 + \frac{1}{n}\sum_{i=1}^{n}X_i\sum_{j=1}^{n}X_j - \frac{2}{n}\sum_{i=1}^{n}X_i\sum_{j=1}^{n}X_j\Big) \\
=& \frac{1}{n-1}\Big(\sum_{i=1}^{n}\mathbb{E}(X_i^2) - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^2) - \frac{1}{n}\sum_{i\neq j}^{n}\mathbb{E}(X_i)\mathbb{E}(X_j)\Big) \\
=& \frac{1}{n-1}\Big(\frac{n-1}{n}\sum_{i=1}^{n}(Var(X_i) + \mathbb{E}(X_i)^2) - (n-1)\mathbb{E}(X_i)^2\Big) \\
=& Var(X_i) + \mathbb{E}(X_i)^2 - \mathbb{E}(X_i)^2 = \sigma^2.
\end{aligned}
$$

As shown in Example A.28, $\mathbb{E}(T_\mu(\boldsymbol{X}_n))$ follows a normal distribution $N(\mu, \frac{\sigma^2}{n})$ and for the sample variance $(n-1)T_{\sigma^2}(\boldsymbol{X}_n)/\sigma^2 \sim \chi^2_{n-1}$ holds - as described in Example A.7. Unfortunately the distribution of $T_{\sigma^2}(\boldsymbol{X}_n)$ is expressed in terms of the unknown parameter σ^2.

A confidence interval is used to determine a likely range for an estimator based on a sample. Hence it is a function, mapping a sample onto a multidimensional interval including the point estimator. We are talking from a confidence interval for a parameter estimate with confidence level $1-\alpha$, if the probability, that the parameter estimate is inside the interval is at least $1-\alpha$.

Example 4.2. ***(Two-sided confidence interval μ - normal distribution $N(\mu, \sigma^2)$)*** *Assume a situation as given in Example 4.1. For the confidence interval we are looking at the term $Z = \frac{T_\mu(\boldsymbol{X}_n)-\mu}{T_\sigma(\boldsymbol{X}_n)/\sqrt{n}}$, which follows, as shown in Example A.9, a Student's-t distribution with $(n-1)$ degrees of freedom. Let $T_{n-1}(x), x \in \mathbb{R}$ denote the cumulative distribution function of Z and $t_{n-1,u} = T^{-1}_{n-1}(u), u \in [0,1]$ the corresponding u-quantile. We are now looking for an interval $C \in CI_{1-\alpha}(\mu) : C = [l, r] : l, r \in \mathbb{R}$:*

$$
\begin{aligned}
P(l \leq \mu \leq r) &= P\Big(\frac{T_\mu(\boldsymbol{X}_n) - r}{T_\sigma(\boldsymbol{X}_n)/\sqrt{n}} \leq \frac{T_\mu(\boldsymbol{X}_n) - \mu}{T_\sigma(\boldsymbol{X}_n)/\sqrt{n}} \leq \frac{T_\mu(\boldsymbol{X}_n) - l}{T_\sigma(\boldsymbol{X}_n)/\sqrt{n}}\Big) \\
&= P(t_{n-1,r_2} \leq Z \leq t_{n-1,l_2}).
\end{aligned}
$$

The last equation holds for some $r_2, l_2 \in \mathbb{R}$, depending on r and l since the inverse of the cumulative distribution function T_{n-1} is a surjective mapping to $\mathbb{R}$. If we want a CI with confidence level $1-\alpha$ the following inequality needs to be satisfied:

$$
P(t_{n-1,r_2} \leq Z \leq t_{n-1,l_2}) = l_2 - r_2 \geq 1 - \alpha. \tag{4.1}
$$

The corresponding CI can be expressed as:

$$
\frac{T_\mu(\boldsymbol{X}_n) - r}{T_\sigma(\boldsymbol{X}_n)/\sqrt{n}} = t_{n-1,r_2} \;\Leftrightarrow\; r = T_\mu(\boldsymbol{X}_n) - t_{n-1,r_2}\frac{T_\sigma(\boldsymbol{X}_n)}{\sqrt{n}}
$$

and

$$\frac{T_\mu(\boldsymbol{X}_n) - l}{T_\sigma(\boldsymbol{X}_n)/\sqrt{n}} = t_{n-1,l_2} \;\Leftrightarrow\; l = T_\mu(\boldsymbol{X}_n) - t_{n-1,l_2}\frac{T_\sigma(\boldsymbol{X}_n)}{\sqrt{n}}.$$

If we now demand (4.1) *as equality and a symmetry in the sense of* $l_2 + r_2 = 1$ *we obtain* $l_2 = 1 - \alpha/2, r_2 = \alpha/2$ *and:*

$$\begin{aligned} C &= [T_\mu(\boldsymbol{X}_n) - t_{n-1,1-\alpha/2}T_\sigma(\boldsymbol{X}_n)/\sqrt{n}, T_\mu(\boldsymbol{X}_n) - t_{n-1,\alpha/2}T_\sigma(\boldsymbol{X}_n)/\sqrt{n}] \\ &= [T_\mu(\boldsymbol{X}_n) - t_{n-1,1-\alpha/2}T_\sigma(\boldsymbol{X}_n)/\sqrt{n}, T_\mu(\boldsymbol{X}_n) + t_{n-1,1-\alpha/2}T_\sigma(\boldsymbol{X}_n)/\sqrt{n}], \end{aligned}$$

since $t_{n-1,a} = -t_{n-1,1-a}$ *holds.*

4.1.1 Two special estimator classes

Very famous approaches of estimator creation are represented by the maximum likelihood and the moment approach. Both approaches have their strengths and weaknesses. The maximum likelihood estimator might be hard to find, since it sometimes cannot be expressed in a computable form, but has great theoretical features. The moment estimator is usually more easy to find as the maximum likelihood estimator but implicates some unfavourable features.

The maximum likelihood estimator is the estimator, for which the maximum likelihood function (probability density function evaluated for the given sample and with distribution parameters as arguments) is maximal. The distribution of a maximum likelihood estimator is approaching a normal distribution for a rising number of observations.

The moment estimator is gained by replacing the theoretical moments through the empirical moments when expressing the distribution parameters as a function of the distribution moments. One of its disadvantages is, that the gained estimators are not necessarily out of the right domain. When starting with a statistical model including the probability family $\mathcal{P} = \{P_\theta | \theta \in \Theta \subseteq \mathbb{R}^d\}$, the moment estimator $\hat{\theta}$ does not necessarily feature $\hat{\theta} \in \Theta$.

4.2 Test Theory

A statistical test can be used to decide between to statements *'The parameter is in the specified range'* and *'The parameter is not in the specified range'*. One of the statements is called null hypothesis H_0, the other is called alternative hypothesis H_1. Formally, a test is defined as function mapping a sample of a statistical model to either 0 or 1. If the mapping equals 0 the sample speaks for H_0 and *we can not reject* H_0; If the mapping equals 1 the sample speaks against H_0 and we reject H_0 for H_1.

A statistical test with significance level $\alpha \in [0, 1]$ rejects H_0 wrongly with a probability smaller or equal to α. This error is called type I error. Thereby we often look at the corresponding p-value - the value for α, for which the specific test evaluated for a sample changes from 0 to 1. This p-value represents the probability to reject H_0 wrongly.

Example 4.3. ***(Comparison of variances - normal distribution)***
Let us assume two iid samples $\boldsymbol{X}_{n_1}$ *with* $X_i \overset{\text{iid}}{\sim} N(\mu_X, \sigma_X^2),\ 1 \leq i \leq n_1$ *and* $\boldsymbol{Y}_{n_2}$ *with* $Y_i \overset{\text{iid}}{\sim} N(\mu_Y, \sigma_Y^2),\ 1 \leq i \leq n_2$*. The parameters* μ_X, μ_Y, σ_X^2 *and* σ_Y^2 *are unknown. We are interested if there is a difference in the variances* σ_X^2 *and* σ_Y^2*. The null hypothesis can be written as* $H_0 : \theta \in \{\theta = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2) | \sigma_X^2 = \sigma_Y^2\}$ *and the alternative hypothesis as* $H_1 : \theta \in \{\theta = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2) | \sigma_X^2 > \sigma_Y^2\}$*. As shown in Example A.11 the quotient*

$$Z = \frac{T_{\sigma^2}(\boldsymbol{X}_{n_1})/\sigma_X^2}{T_{\sigma^2}(\boldsymbol{Y}_{n_2})/\sigma_Y^2}$$

is F_{n_1-1,n_2-1} *distributed. Under* H_0 *the term* Z *can be rewritten as:*

$$Z(\boldsymbol{X}_{n_1}, \boldsymbol{Y}_{n_2}) = \frac{T_{\sigma^2}(\boldsymbol{X}_{n_1})}{T_{\sigma^2}(\boldsymbol{Y}_{n_2})} \sim F_{n_1-1,n_2-1}.$$

If we write $f_{n_1-1,n_2-1,\alpha}$ *for the* α*-quantile of* F_{n_1-1,n_2-1} *the test* ϕ *can be expressed as:*

$$\phi(\boldsymbol{X}_{n_1}, \boldsymbol{Y}_{n_2}) = \begin{cases} 1, & Z(\boldsymbol{X}_{n_1}, \boldsymbol{Y}_{n_2}) \geq f_{n_1-1,n_2-1,1-\alpha} \\ 0, & else \end{cases}.$$

Since $\sigma_Y^2 > \sigma_X^2$ *is not part of* Θ *the numerator of the fraction* Z *should be the one gained from the sample with higher sample variance.*

Another very important test is the Kolmogorov-Smirnov test (Massey Jr, 1951), which is used to compare the sample cumulative distribution function to a reference cumulative distribution function.

Example 4.4. ***(Kolmogorov-Smirnov test)***
We are investigating the null hypothesis $H_0 : \theta \in \Theta_0 = \{\theta_0\}$ *and the alternative hypothesis* $H_1 : \theta \in \Theta_1 = \Theta \setminus \{\theta_0\}$*. Here we could create* Θ *in a way, that a arbitrary high number of distributions are covered (e.g. with a indicator function* $\mathbb{I}$*) and that is why also the cumulative distribution function can be tested. Let* $\boldsymbol{X}_n$ *be the random sample and* F_0 *the continuous cumulative distribution function under* H_0*. The sample cumulative distribution function is given with:*

$$F_{\boldsymbol{X}_n}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_i \leq x),$$

the test variable d *is defined as*

$$d(\boldsymbol{X}_n) = \sup_{x \in \mathbb{R}} |F_0(x) - F_{\boldsymbol{X}_n}(x)|.$$

Under H_0 *the random variable* $d(\boldsymbol{X}_n)$ *is independent of* F_0 *and can be compared to the* $(1-\alpha)$*-quantile* $k_{n,1-\alpha}$ *of the distribution of* $d(\boldsymbol{X}_n)$*, see for example Massey Jr (1951). The test* ϕ *is given with:*

$$\phi(\boldsymbol{X}_n) = \begin{cases} 1, & d(\boldsymbol{X}_n) \geq k_{n,1-\alpha} \\ 0, & else \end{cases}.$$

5. Parameter Regression and Fatigue Data

5.1 Moment Approach

In this chapter we trying to model the structure of fatigue data. We consider situations, where a test object gets stressed periodically with always the same amplitude, called stress-level in this context. The cycles survived by the object are called cycles to failure or time to failure. One is now interested in the distribution of the quantity time to failure depending on the stress-level. Therefore we make the following assumptions:

A1) There is a largest stress-level, denoted here as the minimum stress-level S_{min}, for which the tested object will not suffer failure.

A2) There is a smallest stress-level, denoted here as the maximum stress-level $S_{max} > S_{min}$, for which the tested object will failure immediately.

A3) A higher stress-level reduces the expected time to failure.

A4) A higher stress level reduces the variability for the time to failure.

A5) The distribution of the time to failure is one out of a parametric family of distributions (having k parameters). The parameters depend on the stress-level.

Based on those assumptions we will model the first k central moments in the range between the minimum- and maximum stress-level based on the observations. By using the moment estimators we can subsequently calculate parameter estimators for this stress-range, where possible.
For this approach, we need several observations for multiple stress-level.

5.1.1 Modelling Moments

From the assumptions $A1$ and $A2$ one can conclude, that the expected value for the time to failure is infinity for the minimum stress-level S_{min} and zero for the maximum stress-level S_{max}. In between S_{min} and S_{max} the expectation as a function from the stress-level is strictly monotone decreasing, due to assumption $A3$. Similar considerations lead to strictly monotone decreasing variance function (variance as a function of the stress-level) from the variance at the minimum stress-level denoted as V_{min} to zero at the maximum stress-level. It might be reasonable to assume V_{min} to be infinity - as done here from now on. Suitable functions for modelling the moment structure as function of the stress level could be for example $f(x) = \frac{1}{x}$ or $f(x) = \log(x)$ adapted to minimum and maximum stress level. If there is a need for higher moments more assumptions need to be made on the structure of those higher moments. The empirical moments might be helpful for deciding on the structure.

After deciding on the moment structures we need to adopt the general structure to the observations. We are having k functions $f_1, \cdots, f_k$ which depend on the stress-level and other parameters, like the minimum stress-level or the maximum stress-level. There might also be other parameters which are used to describe the empirical

moments better. The least square or maximum likelihood procedure can now be used to determine the unknown parameters and lead us to the specific moment structure - described by k functions, all of which are only depending on the stress-level.

5.1.2 Example AAW Dataset

In this section and beyond, we are working with the AAW dataset which is taken from Shen (1994, pp. 259 ff.) and can be found in Table 1 listed in Section 1.1. For the distribution family we choose the gamma-family (see Definition A.16), containing two parameters k and θ as used in Section 3.2. The expectation will be modelled with $f(x) = \frac{1}{x}$ and the variance will be modelled with $f(x) = \frac{1}{x^2}$, more precisely we assume for X as time to failure and S as stress-level:

- $E(s, S_{min}, S_{max}, c_1, c_2) = \mathbb{E}[X|S = s] = \big(\frac{S_{max}-S_{min}}{s-S_{min}} - 1\big)c_1$,
- $V(s, S_{min}, S_{max}, c_1, c_2) = Var[X|S = s] = \big(\frac{S_{max}-S_{min}}{s-S_{min}} - 1\big)\frac{S_{max}-S_{min}}{s-S_{min}}c_2$.

In the next step the least square method is used to calculate the parameters S_{min}, S_{max}, c_1 and c_2. Thus we minimize the following target function:

$$\begin{aligned} R(S_{min}, S_{max}, c_1, c_2) = &\sum_{i=1}^{10} \left(\frac{\mu(s_i) - E(s_i, S_{min}, S_{max}, c_1, c_2)}{\mu(s_i)} \right)^2 \\ &+ \sum_{i=1}^{10} \left(\frac{\sigma^2(s_i) - V(s_i, S_{min}, S_{max}, c_1, c_2)}{\sigma^2(s_i)} \right)^2, \end{aligned}$$

whereas $s_i, \quad i \in \{1, \cdots, 10\}$ describe the observation's stress-levels, $\mu(s_i)$ denote the empirical first moment for stress-level s_i and $\sigma^2(s_i)$ describes the central empirical second moment for stress-level s_i. The result can be seen in Table 2 and Figure 5.

Table 2: Parameter estimates - moment structure

Parameter	S_{min}	S_{max}	c_1	c_2
Estimate	49.32	649.36	3038	463480

Estimates of parameters when using the Least Square method for the gamma distribution.

For $X \sim Gamma(k, \theta)$ one can write the expectation as $\mathbb{E}[X] = k \cdot \theta$ and the variance as $Var[X] = k \cdot \theta^2$. This leads us to $\theta(s) = \frac{V(s)}{E(s)} = \frac{c_2}{c_1}\left(\frac{S_{max}-S_{min}}{s-S_{min}}\right)$ and $k(s) = \frac{E^2(s)}{V(s)} = \frac{E(s)}{\theta(s)} = \left(1 - \frac{s-S_{min}}{S_{max}-S_{min}}\right)\frac{c_1^2}{c_2}$. The resulting and observed values are shown in Figure 5.

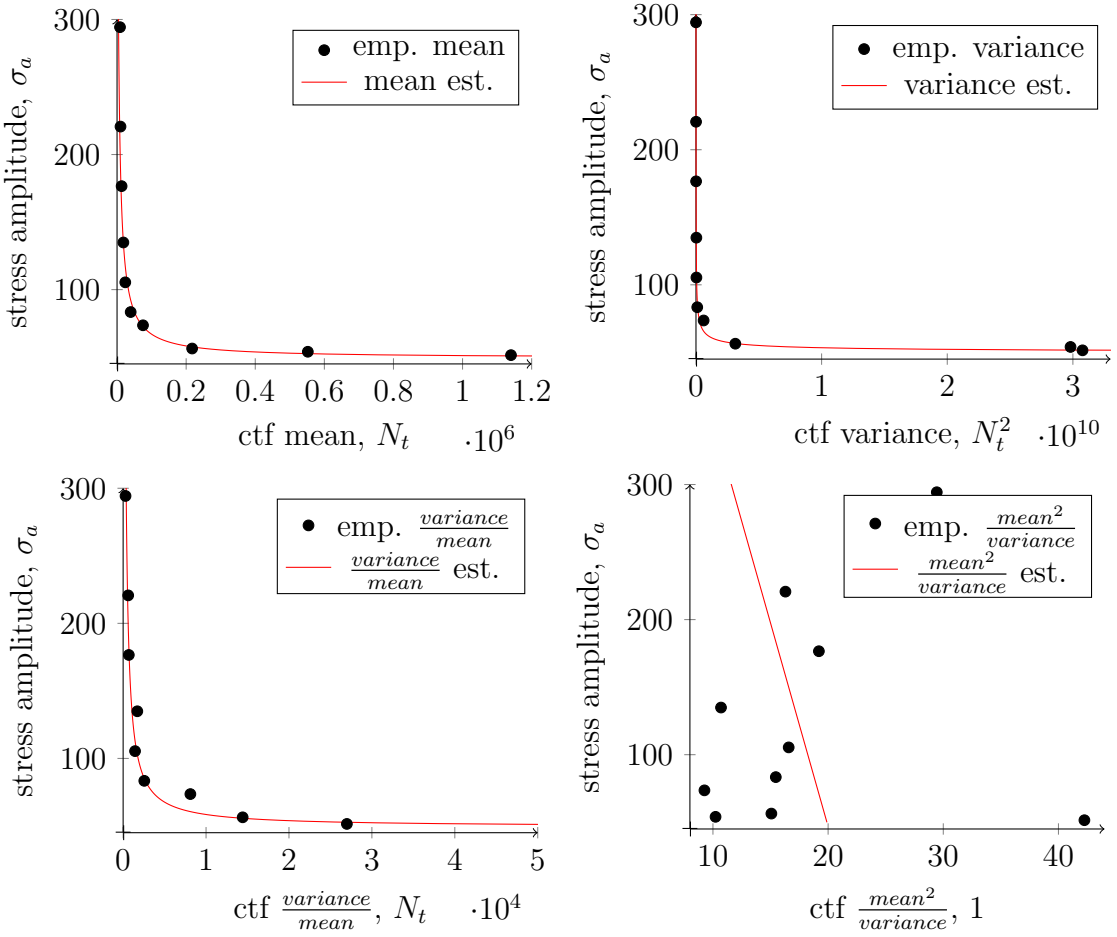


Figure 5: Plot of the empirical moments and the specific moment structure of cycles to failure (ctf) for the first and second central moment using the moment approach according to Section 5.1. Additionally the empirical and calculated values for $\theta(s) = \frac{variance(s)}{mean(s)}$ and $k(s) = \frac{mean(s)^2}{variance(s)}$ are shown.

5.2 Parameter Approach

When apply the above procedure to more complex distributions calculating parameters out of the moments is a nontrivial task. Here, a complex distribution is meaning a distribution, for which the moments can not be written as an easily computable function of the parameters. The Weibull distribution $X \sim Wbl(\lambda, k)$ (see Definition A.15) for example feature the following moments: $\mathbb{E}[X] = \lambda\Gamma(1 + 1/k)$ and $V(X) = \lambda^2\big(\Gamma(1+2/k) - \Gamma^2(1+1/k)\big)$ with $\Gamma(t) = \int_0^\infty x^{t-1}e^x \, dx$ denoting the gamma function. One can consider modelling the parameters instead of the moments to avoid such problems. Similar to modelling the moments one needs to decide on the general structure of the modelling-functions and apply a maximum-likelihood or least square fit to this general structure to get the specific structure of the parameter-functions, called estimated parameter-function in this context.

5.2.1 Example AAW Dataset

Now we try to adopt the Generalized Extreme Value distribution $GEV(\xi, \sigma, \mu), \quad \xi, \mu \in \mathbb{R}, \sigma \in \mathbb{R}_{>0}$ (see Definition A.17) to the given structure. Therefore we first calculate the parameter for each stress-level using the maximum likelihood approach and plot the result, see Figure 6, to get an idea of the functions to use for the modelling. Those parameters will be called observed parameters and be denoted as ξ_i, σ_i and μ_i when the index refers to the stress-level $s_i, \quad i \in \{1, \cdots, 10\}$.

It seems that there is no structure when looking at the maximum likelihood estimates for ξ in Figure 6. This might be an indicator for a random error and suggest to model this parameter as constant. There is also no indication, that ξ depends on the other parameters σ and μ. That is why one can minimize ξ independently from the other parameters and simply take the mean of the ξ-estimates to get the parameter function $\xi(s) = \xi = \frac{1}{10}\sum_{i=1}^{10} \xi_i$.

The structure of the other two parameter looks very similar to those observed in Figure 5 when modelling the first two moments. That is why a similar structure is assumed for the stress-level s and structure parameters $S_{min}, S_{max}, c_1, c_2$:

- $\sigma(s, S_{min}, S_{max}, c_1, c_2) = \big(\frac{S_{max}-S_{min}}{s-S_{min}} - 1\big)c_1$,
- $\mu(s, S_{min}, S_{max}, c_1, c_2) = \big(\frac{S_{max}-S_{min}}{s-S_{min}} - 1\big)c_2$.

The resulting model is denoted as model pr:

$$Y_i \overset{\text{iid}}{\sim} GEV\left(\xi, \left(\frac{S_{max} - S_{min}}{s - S_{min}} - 1\right)c_1, \left(\frac{S_{max} - S_{min}}{s - S_{min}} - 1\right)c_2\right). \tag{5.1}$$

We now gain the estimates $(\hat{\xi}, \hat{S}_{min}, \hat{S}_{max}, \hat{c}_1, \hat{c}_2)$ by applying the maximum likelihood method:

$$(\hat{\xi}, \hat{S}_{min}, \hat{S}_{max}, \hat{c}_1, \hat{c}_2) = \underset{\mathbb{R}^5, S_{min}<S_{max}}{\arg\max} \prod_{i=1}^{n} f(s_i, \xi, S_{min}, S_{max}, c_1, c_2),$$

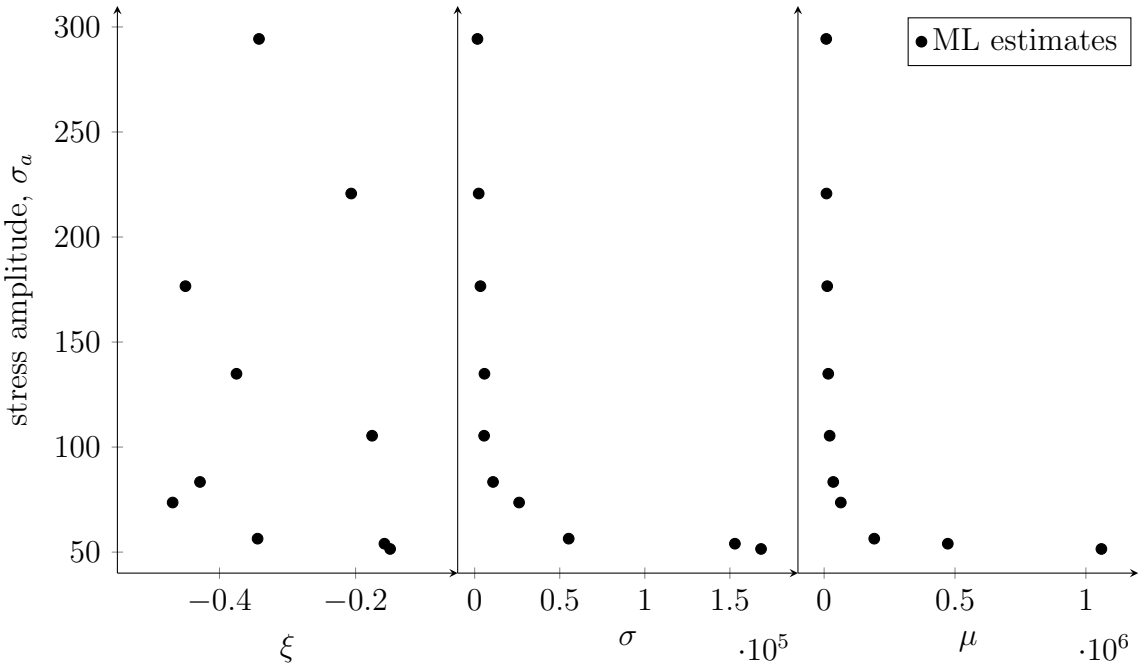


Figure 6: Plot of the GEV maximum likelihood estimates for the AAW dataset at each stress level.

with $f(s_i, \xi, S_{min}, S_{max}, c_1, c_2)$ as the value of the density function given in A.4 with parameters (ξ_f, μ_f, σ_f): $\xi_f = \xi$, $\mu_f = \mu(s, S_{min}, S_{max}, c_1, c_2)$, and $\sigma_f = \sigma(s, S_{min}, S_{max}, c_1, c_2)$. This leads us with computer optimization to the structure parameters shown in Table 3 and Figure 7.

Table 3: Parameter estimates - parameter structure

Parameter	ξ	S_{min}	S_{max}	c_1	c_2
Estimate	-0.16	50.13	21698	22.14	66.64

Estimates of structure parameters when using the maximum likelihood method for the GEV distribution.

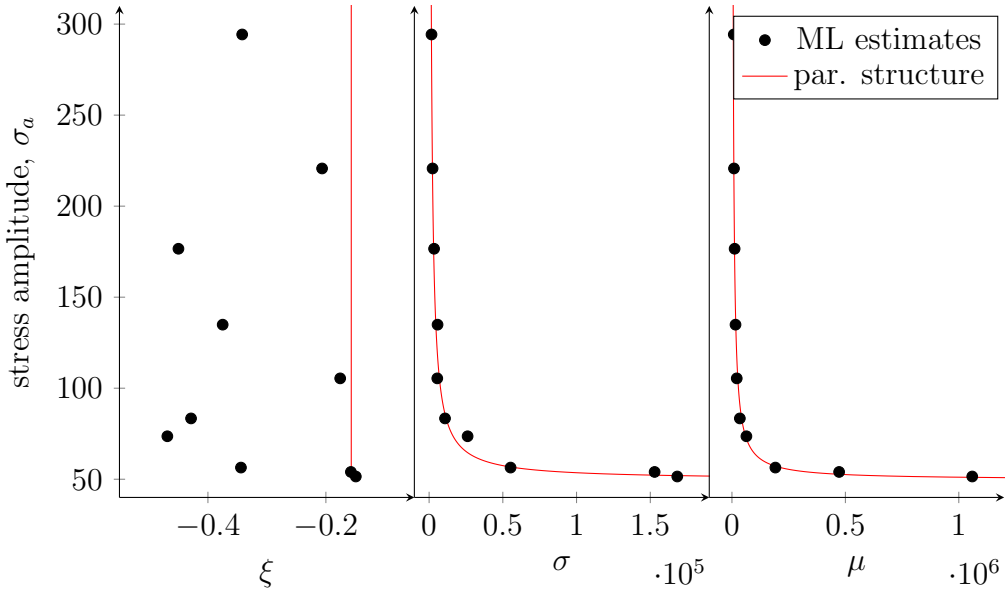


Figure 7: Plot of the GEV maximum likelihood estimates for the AAW dataset at each stress level and the fitted parameter structure.

When now calculating the moments out of this estimated parameter function we get estimated moment functions. Therefore we apply the following relations (we write $\xi(s) = \xi$, $\mu(s) = \mu(s, S_{min}, S_{max}, c_1, c_2)$, and $\sigma(s) = \sigma(s, S_{min}, S_{max}, c_1, c_2)$):

- $E(s) = \mathbb{E}[X|S = s] = \mu(s) + \sigma(s)\frac{\Gamma(1-\xi(s))-1}{\xi(s)}$,
- $V(s) = Var[X|S = s] = \sigma^2(s)\frac{\Gamma(1-2\xi(s))-\Gamma^2(1-\xi(s))}{\xi^2(s)}$,

for $\xi \neq 0, \xi < 1/2$ and for $\xi = 0$:

- $E(s) = \mathbb{E}[X|S = s] = \mu(s) + \sigma(s)\gamma$,
- $V(s) = Var[X|S = s] = \sigma^2(s)\frac{\pi^2}{6}$,

where γ denotes Euler's constant ($\gamma = \lim_{n\to\infty}(-ln(n) + \sum_{k=1}^{n}\frac{1}{k}) \approx 0.577$). For $\xi \geq 1/2$ the variance is infinite - this case is not of interest to us. The results can be seen in Figure 8. The fitted model is visualized in Figure 9.

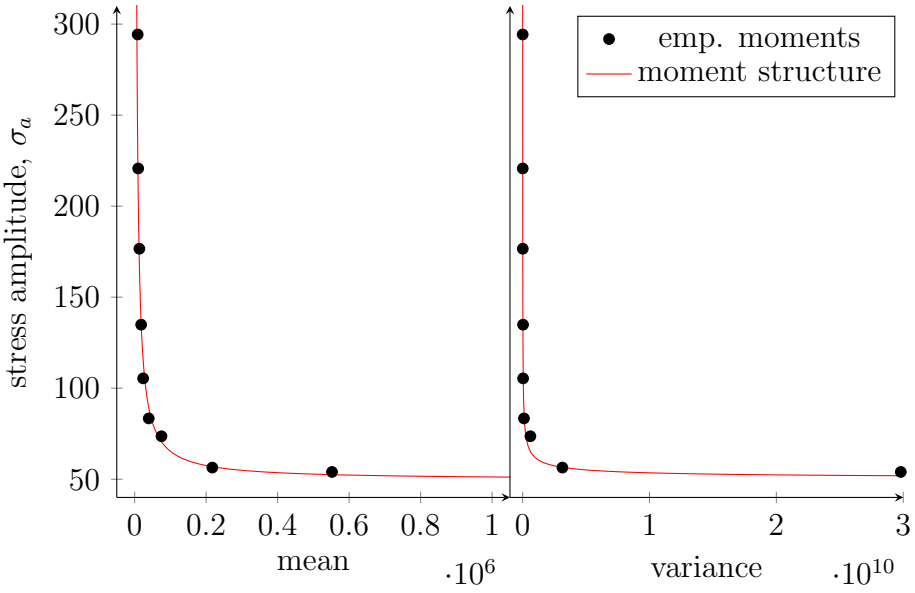


Figure 8: Plot of the GEV empirical moments for the AAW dataset at each stress level and the fitted moment structure.

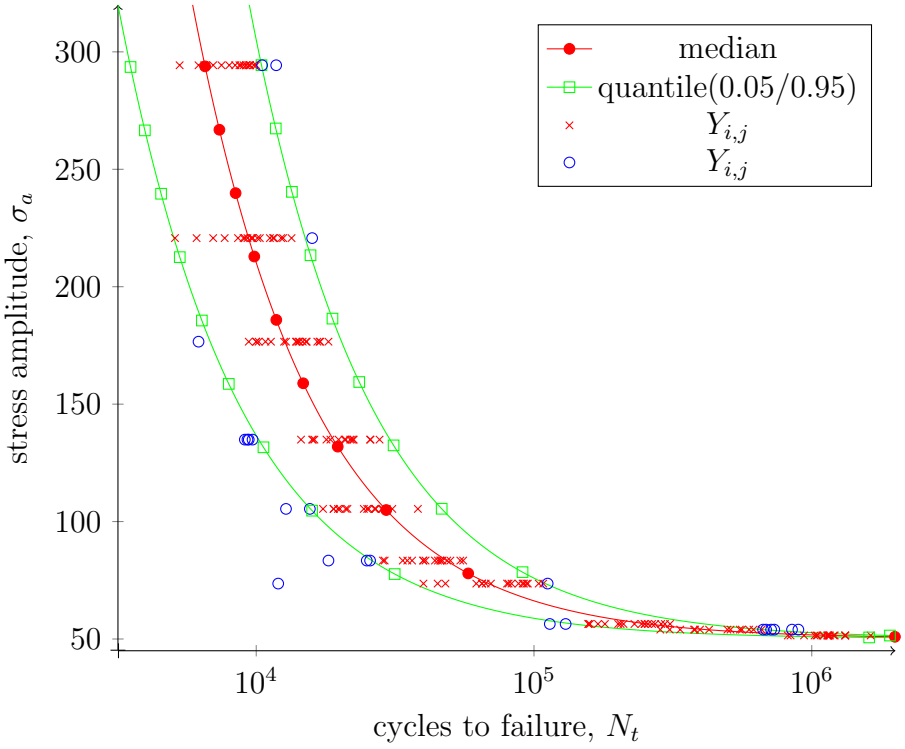


Figure 9: Plot of the GEV estimated median for the AAW dataset at each stress level with quantile lines $(0.05, 0.95)$. The red starts are inside the 0.9 quantile area, the blue circles are outside. There are 88.5% of the observations falling in this area.

5.2.2 Simulation Settings

Simulations can be used to proof the model reliability in a controlled context. These simulations can be used to answer the question: *How adequate can we determine the structure parameters if all assumptions are met?* This is done by defining *'true'* structure parameters and repetitive generating observations based on the base assumptions. In the next step, the generated data is used to calculate the structure parameter estimates. Finally, we can compare the *'true'* structure parameters with their estimates.

There are multiple ways to estimate the structure parameters, we will use the Maximum Likelihood approach and the Least Square approach with different weights. For the Least Square approach, we first calculate the Maximum Likelihood estimates of the distribution parameter vector $\dot{\theta}(x)$ for each stress level x individually. When denoting the structure parameter vector as $\eta = (\eta_1, \cdots, \eta_k)$ we can express the distribution parameter vector $\theta = (\theta_1, \cdots, \theta_j)$ for a given stress level x as function of the structure parameter vector and the stress level $\theta = \theta(\eta, x)$. We gain the ML estimate $\hat{\eta}^{ML}$, the unweighted LS estimate $\hat{\eta}^{LS}_{abs}$, the individual weighted LS estimate $\hat{\eta}^{LS}_{ind}$, and the mean weighted LS estimate $\hat{\eta}^{LS}_{mean}$ by performing the following minimizations:

$$
\begin{aligned}
\hat{\eta}^{ML} &= \underset{\eta' \in \Omega}{\arg\min} \sum_{i=1}^{n} -\log(f_{\theta(\eta', x_i)}(y_i)), \\
\hat{\eta}^{LS}_{abs} &= \underset{\eta' \in \Omega}{\arg\min} \sum_{s=1}^{j} \sum_{i=1}^{n} \left(\theta(\eta', x_i)_s - \dot{\theta}(x_i)_s\right)^2, \\
\hat{\eta}^{LS}_{ind} &= \underset{\eta' \in \Omega}{\arg\min} \sum_{s=1}^{j} \sum_{i=1}^{n} \left(\frac{\theta(\eta', x_i)_s - \dot{\theta}(x_i)_s}{\dot{\theta}(x_i)_s}\right)^2, \\
\hat{\eta}^{LS}_{mean} &= \underset{\eta' \in \Omega}{\arg\min} \sum_{s=1}^{j} \sum_{i=1}^{n} \left(\frac{\theta(\eta', x_i)_s - \dot{\theta}(x_i)_s}{\sum_{x_a = x_i} \dot{\theta}(x_a)_s}\right)^2.
\end{aligned}
\tag{5.2}
$$

In this equations we assume n observations (y_i, x_i) for multiple stress-levels, i.e. values for x_i with multiple occurrences. Those multiple occurrences are needed for the Least Square estimates - for calculating the individual Maximum Likelihood estimates $\dot{\theta}$ at each stress level. The sum over all n observations in the Least square estimates can be reduced to a sum over the stress levels for calculation performance purposes. $f_{\theta(\eta', x)}$ describes the density of the distribution for stress level x and structure parameter vector η'.

We use multiple distributions for the simulation of Y to be able to compare the estimation performance of the approach over different distributions. Besides the Normal- and the Log-normal distribution (see Definition A.1 and A.3) we also take the GEV distribution (see Definition A.17) into account.

The structure parameter vector and its link to the density function parameter vector, the structure functions, for stress x are constructed in line with the considerations of

Table 4: Structure parameter vectors for simulation

Distribution	**Parameter vectors**				
	ξ	S_{min}	S_{max}	c_1	c_2
Normal	-	49.60	346	7261	1401
Log-normal	-	47.30	323	0.092	0.293
GEV	-0.22	49.64	22760	17.25	63.40

Structure parameter vector values used for the simulation for Normal, Log-normal, and GEV distribution.

Section 5.2. They are given for the three distributions $N(\mu_N, \sigma_N), LN(\mu_{LN}, \sigma_{LN})$, and $GEV(\xi_{GEV}, \mu_{GEV}, \sigma_{GEV})$ as:

$$\begin{aligned} \sigma_N = \sigma_{LN} = \sigma_{GEV} &= \sigma(x, S_{min}, S_{max}, c_1, c_2) = \left(\frac{S_{max} - S_{min}}{x - S_{min}} - 1\right) c_1, \\ \mu_N = \mu_{LN} = \mu_{GEV} &= \mu(x, S_{min}, S_{max}, c_1, c_2) = \left(\frac{S_{max} - S_{min}}{x - S_{min}} - 1\right) c_2. \qquad (5.3) \\ \xi_{GEV} &= k. \end{aligned}$$

This leads us to the following structure parameter vectors η:

- Normal distribution: $\eta = \eta_N = (S_{min}, S_{max}, c_1, c_2)^T$,
- Log-normal distribution: $\eta = \eta_{LN} = (S_{min}, S_{max}, c_1, c_2)^T$,
- GEV distribution: $\eta = \eta_{GEV} = (\xi, S_{min}, S_{max}, c_1, c_2)^T$.

We are trying to simulate data similar to the data in Table 1. This means we choose *'true'* structure parameter vectors for the simulation in the range of the parameter estimates gained from this data, see Table 4. Furthermore, we will have ten stress levels and $r \in \{5, 10, 20, 30, 40, 50, 100\}$ observations per stress level. For each distribution and each value of r we generate 10^4 datasets and calculate the structure parameter vector with the Maximum Likelihood method and the three different weighted Least square approaches. An overview of this procedure is given in Table 5.

The simulation procedure was implemented using the free statistic software R. The random number creation is done using the functions *rnorm*, *rlnorm*, and *rgev* (from the library *evir*).

5.2.3 Simulation Results

In this section we present the results of the simulation described in the previous Section 5.2.2. We have generated for each value of r (data points per stress level) and each distribution 10^4 data-sets. Each of these data sets is used to calculate a structure parameter estimation vector. If we now calculate the arithmetic mean of

Table 5: Overview of the simulation procedure

Simulation - Overview
(1. Determination of the distribution (Normal-,Log-normal- or GEV-distribution) and the number of data points per stress level $r \in \{5, 10, 20, 30, 40, 50, 100\}$
(2. Determination of the structure functions (see equations (5.3))
(3. Determination of the *'true'* structure parameter vectors (see Table 4)
(4. Generation of 10^4 times r data points for each of the ten stress levels
(5. Estimation of the structure parameter vectors with four different approaches (see equations (5.2))

Overview of the simulation procedure used to proof the reliability in a controlled context.

these structure parameter estimation vectors up to the k-th estimation we gain a mean estimation vector in dependency of the number of simulations k. The same procedure can be applied to the standard derivation in dependency of the number of simulations k. The mean of the last 1% of the estimation means and standard derivations is used to calculate the values sd and $mean$, leading to the values:

- $mean_r^{dist}$ for $r \in \{5, 10, 20, 30, 40, 50, 100\}$ and $dist \in \{N, LM, GEV\}$,
- sd_r^{dist} for $r \in \{5, 10, 20, 30, 40, 50, 100\}$ and $dist \in \{N, LM, GEV\}$.

Thereby, the abbreviation N denotes estimates gained from the Normal distribution, the abbreviation LN intends the Log-normal distribution estimates, and the abbreviation GEV stands for the GEV-distribution estimates.

Normal distribution

The tables 6, 7, 8, and 9 show the relative error between the true parameter vector and $mean_r^{dist}$ for each vector element in percent. With θ' as vector entry of $mean_r^{dist}$ and θ as the corresponding *'true'* parameter vector entry, it is calculated with:

$$P(\theta') = \left| \frac{\theta - \theta'}{\theta} \right| \cdot 100. \tag{5.4}$$

By comparing Table 6, 7, 8, and 9 we see, that the parameter S_{min} differs not more than one-tenth of a percent to the *'true'* parameter, except for the LS-method, mean weighted, and $r = 5$. The values for the ML-method, Table 6 and the LS-method, individually weighted, Table 8 yield approximately the *'true'* parameters with a deviation of not more than 3.61 or 1.18 percentage for $r = 100$.

Table 6: $P(\theta')$ Normal distribution/ML-method

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	49.6	0.02	0.00	0.01	0.02	0.02	0.02	0.02
S_{max}	346.0	3.10	2.88	3.36	3.41	3.54	3.52	3.61
c_1	7261.0	1.07	1.15	1.43	1.53	1.58	1.61	1.64
c_2	1401.0	3.16	1.63	1.12	0.87	0.76	0.80	0.53

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the Normal distribution/ML method.

Table 7: $P(\theta')$ Normal distribution/LS-method unweighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	49.6	0.02	0.04	0.06	0.09	0.07	0.09	0.09
S_{max}	346.0	331.91	317.23	324.67	358.50	330.72	370.64	383.00
c_1	7261.0	65.35	65.90	67.46	74.22	68.31	74.10	74.29
c_2	1401.0	66.41	66.47	67.65	74.31	68.38	74.21	74.33

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the Normal distribution/LS method - unweighted.

Table 8: $P(\theta')$ Norma distribution/LS-method individually weighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	49.6	1073007.98	0.02	0.01	0.01	0.01	0.00	0.00
S_{max}	346.0	3.10	0.94	0.44	0.26	0.20	0.09	0.07
c_1	7261.0	268663.93	2.37	1.12	0.74	0.58	0.34	0.21
c_2	1401.0	94773.95	12.89	6.02	3.96	3.00	2.31	1.18

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the Normal distribution/LS method - individually weighted.

Table 9: $P(\theta')$ Normal distribution/LS-method mean weighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	49.6	0.34	0.08	0.00	0.08	0.05	0.08	0.09
S_{max}	346.0	278.52	304.44	319.44	347.00	322.69	342.33	338.75
c_1	7261.0	41.34	60.20	68.42	78.03	71.99	78.28	78.20
c_2	1401.0	50.12	62.86	69.50	78.23	72.35	78.43	78.25

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the Normal distribution/LS method - mean weighted.

Log-normal distribution

The tables 10, 11, 12, and 13 show the relative error between the true parameter vector and $mean_r^{dist}$ for each vector element in percent. With θ' as vector entry of $mean_r^{dist}$ and θ as the corresponding *'true'* parameter vector entry, it is calculated with formula 5.4.

Table 10: $P(\theta')$ Log-normal distribution/ML-method

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	47.30	28.85	29.20	29.66	29.99	30.20	30.24	30.18
S_{max}	323.00	2.45	2.48	2.53	2.56	2.59	2.59	2.60
c_1	0.09	206.04	208.82	212.44	214.89	216.62	217.03	216.78
c_2	0.29	150.50	152.98	155.92	157.51	158.72	159.20	159.27

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the Log-normal distribution/ML method.

Table 11: $P(\theta')$ Log-normal distribution/LS-method unweighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	47.30	32.73	25.50	20.05	20.36	23.03	22.76	26.29
S_{max}	323.00	23.55	23.95	24.25	24.22	24.07	24.08	23.86
c_1	0.09	130.44	96.81	71.36	73.02	85.65	84.48	101.15
c_2	0.29	6.87	2.49	2.15	5.53	5.00	6.55	7.15

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the Log-normal distribution/LS method - unweighted.

Table 12: $P(\theta')$ Log-normal distribution/LS-method individually weighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	47.30	2.03	2.09	2.27	2.42	2.47	2.50	2.51
S_{max}	323.00	0.18	0.18	1.53	2.10	2.34	2.55	3.02
c_1	0.09	17.51	17.71	16.68	17.05	17.17	17.08	16.51
c_2	0.29	29.28	13.05	6.10	3.98	3.07	2.32	1.16

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the Log-normal distribution/LS method - individually weighted.

Table 13: $P(\theta')$ Log-normal distributed/LS-method mean weighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	47.30	3.13	3.38	3.80	4.04	4.18	4.29	4.52
S_{max}	323.00	3.18	3.71	4.26	4.59	4.61	4.40	3.86
c_1	0.09	22.64	23.60	25.76	26.88	27.84	28.77	30.97
c_2	0.29	5.99	2.77	1.31	0.93	0.66	0.48	0.26

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the Log-normal distribution/LS method - mean weighted.

The results of the LS-method, mean weighted, Table 13 and LS-method, individually weighted, Table 12 fit the *'true'* parameter best with a maximum deviation of 31 or 16.5 percent for $r = 100$. Thereby, the results of the individually weighted LS-method are more precise as the results of the mean weighted LS-method, except for the parameter c_2.

GEV distribution

The tables 14, 15, 16, and 17 show the relative error between the true parameter vector and $mean_r^{dist}$ for each vector element in percent. With θ' as vector entry of $mean_r^{dist}$ and θ as the corresponding *'true'* parameter vector entry, it is calculated with formula 5.4.

Table 14: $P(\theta')$ GEV-distribution/ML-method

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
ξ	-0.22	11.16	4.98	3.65	2.56	2.05	1.94	1.05
S_{min}	49.64	0.08	0.04	0.02	0.01	0.01	0.01	0.01
S_{max}	22760.00	8.16	8.42	8.53	8.56	8.57	8.57	8.57
c_1	17.25	7.01	8.03	8.92	9.11	9.22	9.32	9.34
c_2	63.40	9.95	9.69	9.65	9.59	9.56	9.55	9.53

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the GEV distribution/ML method.

Table 15: $P(\theta')$ GEV-distribution/LS-method unweighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
ξ	-0.22	11.01	16.40	78.75	16.82	9.79	4.65	4.13
S_{min}	49.64	0.11	0.06	0.04	0.02	0.01	0.01	0.00
S_{max}	22760.00	23.41	16.49	11.21	8.83	7.41	6.36	4.76
c_1	17.25	11.73	9.48	8.68	8.28	8.25	7.69	7.73
c_2	63.40	50.13	26.21	14.78	10.59	8.36	6.56	3.95

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the GEV distribution/LS method - unweighted.

Table 16: $P(\theta')$ GEV-distribution/LS-method individually weighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
ξ	-0.22	77.28	82.90	76.95	63.32	52.09	42.20	17.61
S_{min}	49.64	6.70	0.03	0.25	0.03	0.02	0.02	0.04
S_{max}	22760.00	2.37	11.40	2.86	27.91	32.13	45.41	147.65
c_1	17.25	107.21	96.82	13.51	46.38	31.68	23.84	12.94
c_2	63.40	546.34	154.98	26.46	54.47	35.92	26.07	11.95

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the GEV distribution/LS method - individually weighted.

The results of the ML-method, Table 14 and LS-method, unweighted, Table 15 fit the *'true'* parameter best with a maximum deviation of 9.5 or 7.7 percent for $r = 100$. The LS-method, individually weighted, Table 16 features a deviating trend for parameter S_{max} for increasing values of r. Also the behaviour of parameter S_{max} for the LS-method, mean weighted, Table 17 is surprising - the relative deviation is very

Table 17: $P(\theta')$ GEV-distribution/LS-method mean weighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
ξ	-0.22	55.78	3.61	8.15	5.64	4.59	3.60	2.92
S_{min}	49.64	0.40	0.14	0.08	0.06	0.04	0.04	0.02
S_{max}	22760.00	136.12	94.48	27.93	234.14	122.16	198.63	187.74
c_1	17.25	24.07	26.49	1.15	8.03	3.00	7.25	2.88
c_2	63.40	54.89	43.03	6.11	11.29	3.30	5.38	6.61

'True' parameter values and relative deviation in percent of the simulation results for different values of r and the GEV distribution/LS method - mean weighted.

high (for $r = 20$: 28% and for $r = 30$: 234%). The ML-method also shows a rising trend in the parameter S_{max}'s deviation from the *'true'* parameter; However, this trend is minor and vanishing for $r = 40, 50, 100$.

5.2.4 Simulation Interpretation

In this section we try to answer the question: *Is it possible to reliable estimate the **true** parameters?* The structure parameters are good predictable in some cases, depending on the chosen fitting method (ML-method, LS-method without, individually or mean weighted) and distribution (Normal, Log-normal, or GEV-distribution).

We are introducing another quantity to evaluate the parameter estimates:

$$A(\theta') = \left| \frac{\theta - \theta'}{std(\theta')} \right|, \tag{5.5}$$

with θ' as vector entry of $mean_r^{dist}$, $std(\theta')$ as vector entry of sd_r^{dist}, and θ as the corresponding *'true'* parameter vector entry.

Normal distribution (ML-method preferable)
When looking at Table 8 we see, that the individually weighted LS-method is good compared to the other methods, for high values of r(maximum deviations for $r = 50$: 2.3% and for $r = 100$: 1.2%). Unfortunately, it is useless for small values of r, e.g. for $r = 5$ we gain a deviation of 10^6%. The ML-method, in contrast, is achieving low deviations for high values of r and is also adequate for low values of r. This reliability is the reason why it should be preferred. The tables 19 and 18 show, that the *'true'* parameter is always located within a symmetrical two standard deviation size interval around the parameter estimate.

Log-normal distribution (mean weighted LS-method preferable)
For the Log-normal distribution in combination with the individually weighted LS-method we gain low deviations between the *'true'* and the estimated parameter, but also the standard deviation is very low. This results into a deviation of 19 standard deviations between the parameter estimate of S_{min} and its *'true'* value (see Table 20). For the mean weighted LS-method, which performs tendentially poorer as the

Table 18: Deviation $A(\theta')$ Normal distribution/LS-method individually weighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	49.6	0.0111	0.0220	0.0330	0.0127	0.0294	0.0088	0.0128
S_{max}	346.0	0.1636	0.1573	0.0971	0.0894	0.0723	0.0826	0.0527
c_1	7261.0	0.0102	0.2855	0.1912	0.1637	0.1544	0.1134	0.0855
c_2	1401.0	0.0120	1.0958	0.7569	0.6405	0.5787	0.5293	0.3612

This table shows the deviation of the estimate to the *'true'* parameter, measured in standard deviations of the parameter estimate in case of the Normal distribution and the individually weighted LS-method.

Table 19: Deviation $A(\theta')$ Normal distribution/ML-method

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	49.6	0.0586	0.0139	0.0439	0.0782	0.0905	0.1097	0.1156
S_{max}	346.0	0.1244	0.1259	0.1374	0.1394	0.1427	0.1428	0.1442
c_1	7261.0	0.0954	0.1113	0.1327	0.1423	0.1455	0.1492	0.1505
c_2	1401.0	0.2038	0.1256	0.0934	0.0762	0.0666	0.0717	0.0466

This table shows the deviation of the estimate to the *'true'* parameter, measured in standard deviations of the parameter estimate in case of the Normal distribution and the ML-method.

individually weighted LS-method, this behaviour does not appear (see Table 21). All parameter estimates are not more than 1.15 standard deviations away from the *'true'* parameter values. Hence, it seems, that in this case the mean weighted LS-method is more suitable for the parameter estimation as the individually weighted LS-method.

Table 20: Deviation $A(\theta')$ LogN distribution/LS-method individually weighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	47.300	3.3421	3.1690	4.4813	6.8299	9.4666	12.3368	19.0447
S_{max}	323.000	0.0323	0.0625	0.6618	1.0695	1.2787	1.4264	1.9312
c_1	0.092	2.0071	2.6954	3.4592	4.1471	4.7319	5.0643	6.0463
c_2	0.293	1.8460	1.4498	1.0916	0.9093	0.8241	0.7038	0.5019

This table shows the deviation of the estimate to the *'true'* parameter, measured in standard deviations of the parameter estimate in case of the Log-normal distribution and the individually weighted LS-method.

Table 21: Deviation $A(\theta')$ Log-normal distribution/LS-method mean weighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
S_{min}	47.300	1.1457	1.0357	1.0180	1.0240	1.0324	1.0354	1.0515
S_{max}	323.000	0.5448	0.5880	0.7099	0.8490	0.9133	0.9452	1.0155
c_1	0.092	1.1114	0.9823	0.9485	0.9420	0.9543	0.9738	1.0296
c_2	0.293	0.5537	0.3787	0.2569	0.2286	0.1842	0.1488	0.1146

This table shows the deviation of the estimate to the *'true'* parameter, measured in standard deviations of the parameter estimate in case of the Log-normal distribution and the mean weighted LS-method.

GEV-distribution (unweighted LS-method preferable)
Especially the estimated value of S_{max} is very far away (866 standard deviations) from the *'true'* value when using the ML-method. This is also true, for the parameters c_1 and c_2, although not so extreme (see Table 22). Compared to this deviations, the unweighted LS-method produces much more reliable estimates (see Table 23).

Table 22: Deviation $A(\theta')$ GEV-distribution/ML-method

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
ξ	-0.22	0.2145	0.1551	0.1753	0.1543	0.1460	0.1556	0.1242
S_{min}	49.64	0.1662	0.1195	0.0845	0.0667	0.0636	0.0669	0.0570
S_{max}	22760.00	7.4733	15.7044	39.6367	82.7537	138.5649	322.2551	866.3001
c_1	17.25	0.5395	0.9063	1.4656	1.8293	2.1264	2.4328	3.4927
c_2	63.40	2.0223	2.7179	3.7493	4.4881	5.1152	5.6793	8.0785

This table shows the deviation of the estimate to the *'true'* parameter, measured in standard deviations of the parameter estimate in case of the GEV distribution and the ML-method.

Table 23: Deviation $A(\theta')$ GEV-distribution/LS-method unweighted

Parameter	Real	r=5	r=10	r=20	r=30	r=40	r=50	r=100
ξ	-0.22	0.0007	0.0034	0.0143	0.0151	0.0113	0.0060	0.0078
S_{min}	49.64	0.1216	0.0925	0.0847	0.0521	0.0391	0.0467	0.0108
S_{max}	22760.00	0.8486	0.7732	0.6648	0.5909	0.5568	0.5176	0.4918
c_1	17.25	0.0920	0.4132	1.0552	0.6014	0.2282	1.3181	1.6607
c_2	63.40	0.4191	0.6323	0.6733	0.5383	0.2579	0.4729	0.3895

This table shows the deviation of the estimate to the *'true'* parameter, measured in standard deviations of the parameter estimate in case of the GEV distribution and the unweighted LS-method.

Another question of interest is: *Is there a preferable method of estimating the structure parameter vector?* The answer to this question seems to be no. In our simulation context, it was not possible to clearly identify one estimation technique for all considered distributions and all numbers of r, which yields the most reliable results (see Table 24). For the Normal distribution, we identified the ML-method as most suitable for estimating purposes; The Log-normal distributed data points towards the mean weighted LS-method; The unweighted LS-method and the ML method yield

the best results for the GEV distributed datasets. Also, the individually weighted LS-method seems to be useful in some contexts.

Table 24: Comparison - best methods $r = 20$

Parameter	F(θ')	Normal ML	Log-normal LS - mean weighted	GEV LS - unweighted	GEV ML
ξ	$\left\lvert\frac{\theta-\theta'}{std(\theta')}\right\rvert$	-	-	0.0143	0.1753
	$\left\lvert\frac{std(\hat{\theta})}{\theta}\right\rvert$	-	-	30.7579	0.2011
S_{min}	$\left\lvert\frac{\theta-\theta'}{std(\theta')}\right\rvert$	0.0439	1.0180	0.0847	0.0845
	$\left\lvert\frac{std(\hat{\theta})}{\theta}\right\rvert$	0.0025	0.0063	0.0044	0.0022
S_{max}	$\left\lvert\frac{\lvert\theta-\theta'\rvert}{std(\theta')}\right\rvert$	0.1374	0.7099	0.6648	39.6367
	$\left\lvert\frac{std(\hat{\theta})}{\theta}\right\rvert$	0.2365	0.5272	0.1899	0.0024
c_1	$\left\lvert\frac{\theta-\theta'}{std(\theta')}\right\rvert$	0.1327	0.9485	1.0552	1.4656
	$\left\lvert\frac{std(\hat{\theta})}{\theta}\right\rvert$	0.1094	1.1985	0.0757	0.0559
c_2	$\left\lvert\frac{\theta-\theta'}{std(\theta')}\right\rvert$	0.0934	0.2569	0.6733	3.7493
	$\left\lvert\frac{std(\hat{\theta})}{\theta}\right\rvert$	0.1215	1.1382	0.1912	0.0235

Comparison of the best methods for the Normal, Log-normal, and GEV distribution simulation results and $r = 20$. We used the two quantities **deviation from *'true'* parameter in standard deviations** $\left\lvert\frac{\theta-\theta'}{std(\theta')}\right\rvert$ and **scaled standard deviation** $\left\lvert\frac{std(\theta')}{\theta}\right\rvert$.

5.2.5 Further considerations

Next steps in the analysis would be investigating the effects of structure parameter fluctuation on the distribution parameters for the different stress levels as well as on their moments. This would enable us to compare the methods independent of the chosen distribution.

A variation of more parameters included in the simulation would allow further conclusions on the procedure. We could, for example, vary the number of stress levels, as well as the location of the stress levels. How would these changes influence data points and predictions within the stress level scope $[\min\{X_i\}, \max\{X_i\}]$? Also, the implementation of the optimization allows multiple settings; How do they influence the results?

The procedure is not representative for the general fitting approach since it was applied to exactly one structure parameter vector. Universal statements and conclusions are not possible.

6. Standardizing Approach

One of the biggest problems when working with fatigue data is the small size of data samples for evaluation, due to high testing costs. That is why the possibilities to generate a prediction based on the observations, e.g. to calculate the mean and variance, is limited. Therefore additional assumptions have to be made, such as the specification of the underlying distribution. In this section, an approach for calculating cycle thresholds that will be reached by a certain share of stressed components before failure will be presented. For describing this threshold we introduce the term *quantile*.

6.1 Quantiles

Definition 6.1. ***(quantile)***
The unique α-quantile of a random variable X with probability distribution function $F_X(x)$ and $\alpha \in [0,1]$ is defined as:

$$q_\alpha^X = F_X^{\leftarrow}(\alpha) := \inf\{x : F_X(x) \geq \alpha\}.$$

Within a given context the random variable can be omitted and the α-quantile can be written as q_α.

Remark 6.2. *For random variables with continuous and strictly monotone increasing probability distribution function $F(x)$ the α-quantile can be written as $q_\alpha = \inf\{x|F(x) \geq \alpha\} = F^{-1}(\alpha)$.*

Example 6.3. ***(empirical quantile)***
Considering an independent and identically distributed data sample $\{X_i|1 \leq i \leq n\}$ one can calculate quantile estimators $\hat{q}$ by evaluating the empirical distribution function $F_n(x)$ given as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{\{x \geq x_i\}},$$

for x_i as realisation of the random variable X_i. This leads to the empirical quantile and hence to the quantile estimator $\hat{q}_\alpha$

$$\hat{q}_\alpha = q_\alpha^n = F_n^{\leftarrow}(\alpha) = \inf\{x : F_n(x) \geq \alpha\}.$$

If we now sort the observations and denote the i-th smallest observation with $x_{[i]}$ we observe, that only $n+1$ different values for $F_n(x)$ are taken on: $0, F_n(x_{[1]}) = \frac{1}{n}, \cdots, F_n(x_{[n]}) = \frac{n}{n} = 1$ for each of which $x_{[i]} = inf\{x : F_n(x) = F_n(x_{[i]})\}$. This especially leads to n possible values for q_α^n ($\alpha > 0$): $x_{[1]}, \cdots, x_{[n]}$ and simplifies the term q_α^n to:

$$\begin{aligned} q_\alpha^n &= \inf\{x : F_n(x) \geq \alpha\} \\ &= \min_{1 \leq i \leq n}\{x_{[i]} : F_n(x_{[i]}) \geq \alpha\} \\ &= \min_{1 \leq i \leq n}\left\{x_{[i]} : \frac{i}{n} \geq \alpha\right\} \\ &= \min_{1 \leq i \leq n}\{x_{[i]} : i \geq \alpha \cdot n\} = x_{[\lceil \alpha \cdot n \rceil]}. \end{aligned}$$

Remark 6.4. *The empirical quantile is a distribution free approach, since no assumption on the random variable distribution is needed. It is limited by the number of observations to generate this estimation of the quantile, e.g. for α values in the interval $[0, 1/n]$ the quantile estimation will always be $x_{[1]}$.*

Example 6.5. ***(empirical quantile - confidence interval)***
For calculating a confidence interval for the α-quantile associated with an iid random sample $\{X_i | 1 \leq i \leq n\}$ we study the random variable $Y = \sum_{i=1}^{n} Y_i$ with $Y_i = \mathbf{I}_{\{q_\alpha \geq X_i\}}$. Since Y_i is a Bernoulli random variable with parameter α and the variables Y_i are independent and identically distributed we conclude, that Y is a $Bin(n, p)$ random variable, a Binomial distributed random variable with parameters n and p. Furthermore we can see a connection to the empirical density function $F_n(x)$ of the random sample:

$$\begin{aligned} F_n(q_\alpha) &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{I}_{\{q_\alpha \geq X_i\}} \\ &= \frac{1}{n} \sum_{i=1}^{n} Y_i = \frac{1}{n} Y. \end{aligned}$$

We can now determine a confidence interval $[c_l^Y, c_u^Y]$ for the $Bin(n, p)$ distributed random variable Y and use it to generate a confidence interval $[c_l^{q_\alpha}, c_u^{q_\alpha}]$ for the α-quantile, given with:

$$\begin{aligned} c_l^{q_\alpha} &= F_n^{\leftarrow}(c_l^Y / n) = x_{\left[\lceil c_l^Y \rceil\right]}, \\ c_u^{q_\alpha} &= F_n^{\leftarrow}(c_u^Y / n) = x_{[\lceil c_u^Y \rceil]}. \end{aligned}$$

This can be proved by the following equation:

$$\begin{aligned} 1 - \alpha = \mathbb{P}[c_l^Y \leq Y \leq c_u^Y] &= \mathbb{P}[c_l^Y \leq n \cdot F_n(q_\alpha) \leq c_u^Y] \\ &= \mathbb{P}[F_n^{\leftarrow}(c_l^Y / n) \leq q_\alpha \leq F_n^{\leftarrow}(c_u^Y / n)] \\ &= \mathbb{P}[c_l^{q_\alpha} \leq q_\alpha \leq c_u^{q_\alpha}]. \end{aligned}$$

Especially when working with quantiles close to zero ($\alpha \approx 1/n$) and one ($\alpha \approx (n-1)/n$) using the empirical quantiles is not suitable. Therefore one can assume an underlying distribution and work with it's quantiles.

Example 6.6. ***(quantile - underlying distribution)***
For an independent and identically distributed data sample of random variables $\{X_i | 1 \leq i \leq n\}$ and there realisation $\{x_i | 1 \leq i \leq n\}$ one can assume an underlying distribution, e.g. the normal distribution $X_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. In a first step the parameter estimators $\hat{\mu}$ and $\hat{\sigma}^2$ for this sample are calculated by using for example the maximum likelihood method:

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^{n} x_i \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2. \end{aligned}$$

Afterwards the quantile function of the distribution is determined. In case of the strictly monotone probability distribution function of the normal distribution one can use the inverse of the distribution function of the standard normal distribution $\Phi(x)$ to get the quantile of the parameter μ:

$$\hat{q}_\alpha(\mu) = \hat{\mu} + \hat{\sigma}\Phi^{-1}(\alpha).$$

In general an approximative confidence interval with confidence level $1-\tau$ for $\tau \in [0,1]$ can be constructed for $\hat{q}_\alpha$ by using the asymptotic distribution of maximum likelihood estimates θ^{ML}. For this approach the standard deviation of the parameter estimates $sd(\theta_i^{ML})$ are calculated and used to derive the asymptotic confidence intervals

$$[\theta_i^{ML} + sd(\theta_i^{ML})\Phi^{-1}(\tau/(2k)), \theta_i^{ML} + sd(\theta_i^{ML})\Phi^{-1}(1-\tau/(2k))], \quad 1 \leq i \leq k$$

with confidence level $1-\tau/k$, whereas k denotes the number of estimated parameters and θ_i^{ML} denotes the maximum likelihood estimate of the i-th parameter. The k-dimensional interval Θ constructed out of the parameter confidence intervals is then a confidence interval with confidence level $1-\tau$ for the parameter vector. As a final step one has to determine those two parameter-sets out of Θ, which yield the smallest and the biggest value for the quantile of the random variable X; They describe the confidence interval I_τ for the quantile. If $q(\theta)_\alpha$ describes the α quantile of the random variable X, when evaluated for the parameter-vector $\theta \in \Theta$, the confidence interval I_τ for q_α can be written as

$$I_\tau = [a, b],$$

with

$$a = \inf_{\theta \in \Theta}\{q(\theta)_\alpha\},$$
$$b = \sup_{\theta \in \Theta}\{q(\theta)_\alpha\}.$$

In the special case, where the distribution of the parameter estimates is known this distribution can be used to construct a exact confidence interval for the quantile.
For our example we gain $\hat{\mu} \sim N(\mu, \sigma^2/n)$ and $\hat{\sigma}^2 \sim \frac{\sigma^2}{n}\chi^2_{n-1}$. Since $\hat{q}_\alpha$ is monotone increasing with $\hat{\mu}$ and monotone decreasing with $\hat{\sigma}$ (for $\alpha < 0.5$) we have to calculate the confidence interval $[a_\mu, b_\mu]$ and $[a_\sigma, b_\sigma]$ with confidence level $1-\tau/2$ to get the confidence interval I_τ for q with confidence level $1-\tau$ given with

$$I_\tau = [a_\mu + b_\sigma\Phi^{-1}(\alpha), b_\mu - b_\sigma\Phi^{-1}(\alpha)].$$

6.2 Terminology and Base Assumptions

In this section further on we will work with fatigue data consisting of observations $(X_i,Y_{i,j})$, whereas X_i describes the stress level under which the fatigue test was carried out and can be seen as constant combined with $Y_{i,j}$ describing the number of cycles to failure. Moreover we assume $i \in \{1, \cdots, n\}$ and $j \in \{1, \cdots, k\}$. Additionally we assume independence and identically distributed variables $Y_{i,j} \overset{\text{iid}}{\sim} F_i'$ at each stress level X_i and independence of all observations $Y_{i,j}$ for distributions F_i'.

6.3 Main Idea

The main idea of the following approach is, to map the observations to gain independent and identically distributed observations. If such a function $T(y, \omega) : \mathbb{R}_{>0} \times \Omega \to \mathbb{R}$, strictly monotone in x, exists we automatically gain

$$Z_{i,j} = T(Y_{i,j}, \omega_i) \overset{\text{iid}}{\sim} F_i, \quad \forall i, j$$

for a distribution F_i, because of the independence assumed at each stress level ($Y_{i,j} \overset{\text{iid}}{\sim} F_i'$); Additionally we assume

$$F_i = F, \quad \forall i$$

for a distribution F. $\omega_i \in \Omega$ is a quantity dependent on the mapping and contains information about the stress-level associated with the first argument of the function.

Example 6.7. ***(Normal distribution)***
If we assume normally distributed fatigue data, i.e. we assume $Y_{i,j} \sim N(\mu_i, \sigma_i^2)$ with $\mu_i \neq \mu_j$ and $\sigma_i^2 \neq \sigma_j^2$ for $i \neq j$ the mapping $T'(y, (\mu, \sigma)) : \mathbb{R}_{>0} \times (\mathbb{R} \times \mathbb{R}_{>0}) \to \mathbb{R}$ could be

$$Z_{i,j}' = T'(Y_{i,j}, (\mu_i, \sigma_i)) = \frac{Y_{i,j} - \mu_i}{\sigma_i}.$$

This would lead to $Z_{i,j}' \sim N(0, 1)$. Since in general μ_i and σ_i are unknown one could work with the mapping

$$Z_{i,j} = T(Y_{i,j}, \{Y_{i,l} | 1 \leq l \leq k\}) = \frac{Y_{i,j} - \frac{1}{n}\sum_{l=1}^{n} Y_{i,l}}{\sqrt{\frac{1}{n-1}\sum_{l=1}^{n}(Y_{i,l} - \frac{1}{n}\sum_{s=1}^{n} Y_{i,s})^2}} = \frac{Y_{i,j} - \hat{\mu}_i}{\hat{\sigma}_i},$$

whereas $\hat{\mu}_i$ and $\hat{\sigma}_i$ describe estimates for μ_i and σ_i leading to

$$Z_{i,j} \sim T_{k-1},$$

a student's t distribution, as described in Section A.1. In this situation ω_i is the tuple $(\hat{\mu}_i, \hat{\sigma}_i)$.

This approach enables us to use $n \cdot k$ observations for estimating quantiles after the transformation. In the next step, we will re-transform the estimated quantiles back to the original axis. Therefore the inverse of the function $T(y, \omega)$ regarding the first argument y by fixed second argument ω is needed

Definition 6.8. ***(transformation/re-transformation)***
A transformation function is a function $T_\omega(y) = T(y, \omega) : \mathbb{R}_{>0} \times \Omega \to \mathbb{R}$, which is continuous and strictly monotone in x when having ω fixed:

$$\forall \omega \in \Omega : \forall y_1, y_2 \in \mathbb{R} : y_1 < y_2 \Rightarrow T_\omega(y_1) < T_\omega(y_2),$$

or

$$\forall \omega \in \Omega : \forall y_1, y_2 \in \mathbb{R} : y_1 < y_2 \Rightarrow T_\omega(y_1) > T_\omega(y_1).$$

The corresponding re-transformation function $T_\omega^{-1}(z) = T^{-1}(z,\omega) : \mathbb{R} \times \Omega \to \mathbb{R}$ *for argument* z *and* ω *is defined as the only solution* $y \in \mathbb{R}$ *to the equation*

$$T_\omega(y) = z.$$

If $\forall z \in \mathbb{R}, \forall \omega \in \Omega : T_\omega^{-1}(z) > 0$ *holds, the transformation and re-transformation functions are denoted as positive transformation and re-transformation functions.*

Remark 6.9. *Since the quantity cycles to failure is a strictly non negative quantity we are interested in providing positive estimates for it. For positive transformation and re-transformation functions, this is automatically true.*

Example 6.10. ***(Normal distribution - re-transformation function)***
If we now consider the situation in Example 6.7, we obtain the re-transformation function $T_\omega^{-1}(z)$ *as*

$$T_{(\hat{\mu}_i,\hat{\sigma}_i)}^{-1}(z) = z \cdot \hat{\sigma}_i + \hat{\mu}_i,$$

since

$$T_{(\hat{\mu}_i,\hat{\sigma}_i)}(z \cdot \hat{\sigma}_i + \hat{\mu}_i) = \frac{(z \cdot \hat{\sigma}_i + \hat{\mu}_i) - \hat{\mu}_i}{\hat{\sigma}_i} = z$$

holds. For $z < -\frac{\hat{\mu}_i}{\hat{\sigma}_i}$ *the value for* $T_{(\hat{\mu}_i,\hat{\sigma}_i)}^{-1}(z)$ *is negative. Therefore the transformation and re-transformation functions are not positive.*

6.4 Estimating quantiles

Estimating quantiles $\hat{q}_\alpha$ and their confidence intervals $[q_\alpha^l, q_\alpha^u]$ for an independent and equally distributed dataset $Z_{i,j} = T_{\omega_i}(Y_{i,j}), \quad i \in \{1, \cdots, n\}, j \in \{1, \cdots, k\}$ seems to be easily possible considering Section 6.1. We now have to re-transform these quantiles back using the re-transform function $T_{\omega_i}^{-1}(z)$. For deterministic re-transformation functions this can be easily done.

Proposition 6.11. ***(re-transformation of quantiles)***
Let Z *be a random variable with continuous and strictly monotone increasing probability distribution function,* q_α *the corresponding* α*-quantile and* $[a, b]$ *a confidence interval for a quantile estimator* $\hat{q}_\alpha$ *with confidence level* $1 - \tau$ *for* $\tau \in [0, 1]$*. For the random variable* $Y = T_\omega^{-1}(Z)$ *with* $T_\omega^{-1}(x)$ *as re-transformation function (with* $T_\omega(x)$ *as deterministic transformation function for deterministic values of* ω*) the following holds:*

a) $T_\omega^{-1}(q_\alpha)$ *is the* α*-quantile for* Y*, if* $T_\omega(x)$ *is monotone increasing, otherwise it is the* $(1 - \alpha)$*-quantile for* Y*.*

b) The interval $[T_\omega^{-1}(a), T_\omega^{-1}(b)]$ *for a monotone increasing re-transformation function and* $[T_\omega^{-1}(b), T_\omega^{-1}(a)]$ *for a monotone decreasing re-transformation is a confidence interval for the* α*-quantile* $T_\omega^{-1}(q_\alpha)$ *for* Y *with confidence level* $1 - \tau$*.*

Proof. **a)**
Since we are in the situation of a random variable Z with a continuous and strictly

monotone increasing probability distribution function $F_Z(x)$ the α-quantile can be rewritten as:

$$q_\alpha^Z = \inf\{x | F_Z(x) \geq \alpha\} = F_Z^{-1}(\alpha).$$

Since $T_\omega(x)$ is a strictly monotone function in x, also $T_\omega^{-1}(x)$ is strictly monotone. We can now show, that also $F_{T_\omega^{-1}(Z)} = F_Y$ strictly monotone. It is continuous since $T_\omega(x)$ is continuous.

Assumption: There is an interval $[a, b] : a \neq b \wedge \mathbb{P}[a < Y \leq b] = F_Y(b) - F_Y(a) = 0$. For a, b out of F_Y's domain of definition $T_\omega(a)$ and $T_\omega(b)$ are out of F_Z's domain of definition with $T_\omega(a) \neq T_\omega(b)$, since T_ω is strictly monotone. Therefore we can write for monotone increasing transformation function $T_\omega(x)$:

$$0 < \mathbb{P}[T_\omega(a) < Z \leq T_\omega(b)] = \mathbb{P}[a < T_\omega^{-1}(Z) = Y \leq b] = F_Y(b) - F_Y(a) = 0,$$

and for monotone decreasing transformation function $T_\omega(x)$:

$$0 < \mathbb{P}[T_\omega(a) < Z \leq T_\omega(b)] = \mathbb{P}[a > T_\omega^{-1}(Z) = Y \geq b] = F_Y(b) - F_Y(a) = 0.$$

This now implies, that $q_\alpha^Y = F_Y^{-1}(\alpha)$ holds and we finally obtain:

$$\alpha = \mathbb{P}[Z \leq q_\alpha] = \begin{cases} \mathbb{P}[Y \leq T_\omega^{-1}(q_\alpha)] = F_Y(T_\omega^{-1}(q_\alpha)), & T_\omega^{-1}(x) \text{ mon. increasing} \\ \mathbb{P}[Y \geq T_\omega^{-1}(q_\alpha)] = 1 - F_Y(T_\omega^{-1}(q_\alpha)), & T_\omega^{-1}(x) \text{ mon. decreasing} \end{cases}.$$

b)

$$1 - \tau \leq \mathbb{P}[a \leq q_\alpha \leq b] = \begin{cases} \mathbb{P}[T_\omega^{-1}(a) \leq T_\omega^{-1}(q_\alpha) \leq T_\omega^{-1}(b)], & T_\omega^{-1}(x) \text{ mon. increasing} \\ \mathbb{P}[T_\omega^{-1}(b) \leq T_\omega^{-1}(q_\alpha) \leq T_\omega^{-1}(a)], & T_\omega^{-1}(x) \text{ mon. decreasing} \end{cases}.$$

□

When applying the procedure of Proposition 6.11 for ω we would not get the α-quantile, but a point estimator for the α-quantile, since we are working with the point estimator for ω. If we use the appropriate quantiles of ω instead of the point estimator ω we are able to construct the desired α-quantile. Getting such a quantile might be a problem. This is expressed in Proposition 6.12.

Proposition 6.12. ***(stochastic re-transformation of quantiles)***
Let Z be a random variable with continuous and strictly monotone increasing probability distribution function, $[a, b]$ a confidence interval for the α-quantile q_α with confidence level $1 - \tau_1$ for $\tau_1 \in (0, 0.5)$ and C a confidence area for $\omega \in \Omega$ with confidence level $1 - \tau_2$ for $\tau_2 \in (0, 0.5)$. For the random variable $Y = T_\omega^{-1}(Z)$ with $T_\omega^{-1}(x)$ as re-transformation function (with $T_\omega(x)$ as transformation function and stochastic values for ω) the following holds:

- *The interval $[T_{\omega^1}^{-1}(a), T_{\omega^2}^{-1}(b)]$ for a monotone increasing re-transformation function and $[T_{\omega^3}^{-1}(b), T_{\omega^4}^{-1}(a)]$ for a monotone decreasing re-transformation is a confidence interval for the α-quantile $T_\omega^{-1}(q_\alpha)$ for Y with confidence level $1 - (\tau_1 +$*

τ_2), *whereas* $\omega^i, i \in \{1,2,3,4\}$ *is given with:*

$$\begin{aligned}
\omega^1 &= \arg\inf_{\omega \in C} T_\omega^{-1}(a), \\
\omega^2 &= \arg\sup_{\omega \in C} T_\omega^{-1}(b), \\
\omega^3 &= \arg\inf_{\omega \in C} T_\omega^{-1}(b), \\
\omega^4 &= \arg\sup_{\omega \in C} T_\omega^{-1}(a).
\end{aligned}$$

Proof.
We only proof the statement for a monotone increasing function T_ω:

$$\begin{aligned}
\mathbb{P}[T_{\omega^1}^{-1}(a) \leq T_\omega^{-1}(q_\alpha) \leq T_{\omega^2}^{-1}(b)] &= 1 - \mathbb{P}[T_\omega^{-1}(q_\alpha) \notin [T_{\omega^1}^{-1}(a), T_{\omega^2}^{-1}(b)] \wedge \omega \in \Omega] \\
&= 1 - \mathbb{P}[T_\omega^{-1}(q_\alpha) \notin [T_{\omega^1}^{-1}(a), T_{\omega^2}^{-1}(b)] \wedge \omega \in C] \\
&\quad - \mathbb{P}[T_\omega^{-1}(q_\alpha) \notin [T_{\omega^1}^{-1}(a), T_{\omega^2}^{-1}(b)] \wedge \omega \in C^c] \\
&\geq 1 - \mathbb{P}[T_\omega^{-1}(q_\alpha) \notin [T_\omega^{-1}(a), T_\omega^{-1}(b)] \wedge \omega \in C] \\
&\quad - \mathbb{P}[T_\omega^{-1}(q_\alpha) \notin [T_{\omega^1}^{-1}(a), T_{\omega^2}^{-1}(b)] \wedge \omega \in C^c] \\
&\geq 1 - \mathbb{P}[T_\omega^{-1}(q_\alpha) \notin [T_\omega^{-1}(a), T_\omega^{-1}(b)]] - \mathbb{P}[\omega \in C^c] \\
&= 1 - \mathbb{P}[q_\alpha \notin [a,b]] - \mathbb{P}[\omega \in C^c] \\
&\geq 1 - \tau_1 - \tau_2 = 1 - (\tau_1 + \tau_2)
\end{aligned}$$

□

6.5 The AAW Dataset

Now we apply the described approach to the AAW-Dataset, which can be found in Table 1 to gain the model *sa*. The transformation function is the following:

$$z_{i,j} := T_{\omega_i}(y_{i,j}) = \log\left(e^{\frac{v_{i,j}-\omega_i^1}{\omega_i^2}} + 1\right), \quad \text{for} \quad v_{i,j} = \frac{1}{y_{i,j}}, \tag{6.1}$$

with

$$\omega_i = (\omega_i^1, \omega_i^2) = (\bar{v}_i, s_{v_i}) = \left(\frac{1}{k}\sum_{j=1}^{k} v_{i,j}, \sqrt{\frac{1}{k-1}\sum_{j=1}^{k}\left(v_{i,j} - \bar{v}_i\right)^2}\right).$$

As distribution we assume the two parameter Frechet distribution (e.g. see (Franke et al., 2015)) with cumulative distribution function $F_{b,s}(x)$, which can be written as:

$$F(x) = e^{-(b/x)^s}, \quad b > 0, s > 0, x > 0.$$

Hence the model is given with:

$$z_{i,j} = \log\left(e^{\frac{v_{i,j}-\bar{v}_i}{s_{v_i}}} + 1\right) \overset{\text{iid}}{\sim} F_{b,s} \tag{6.2}$$

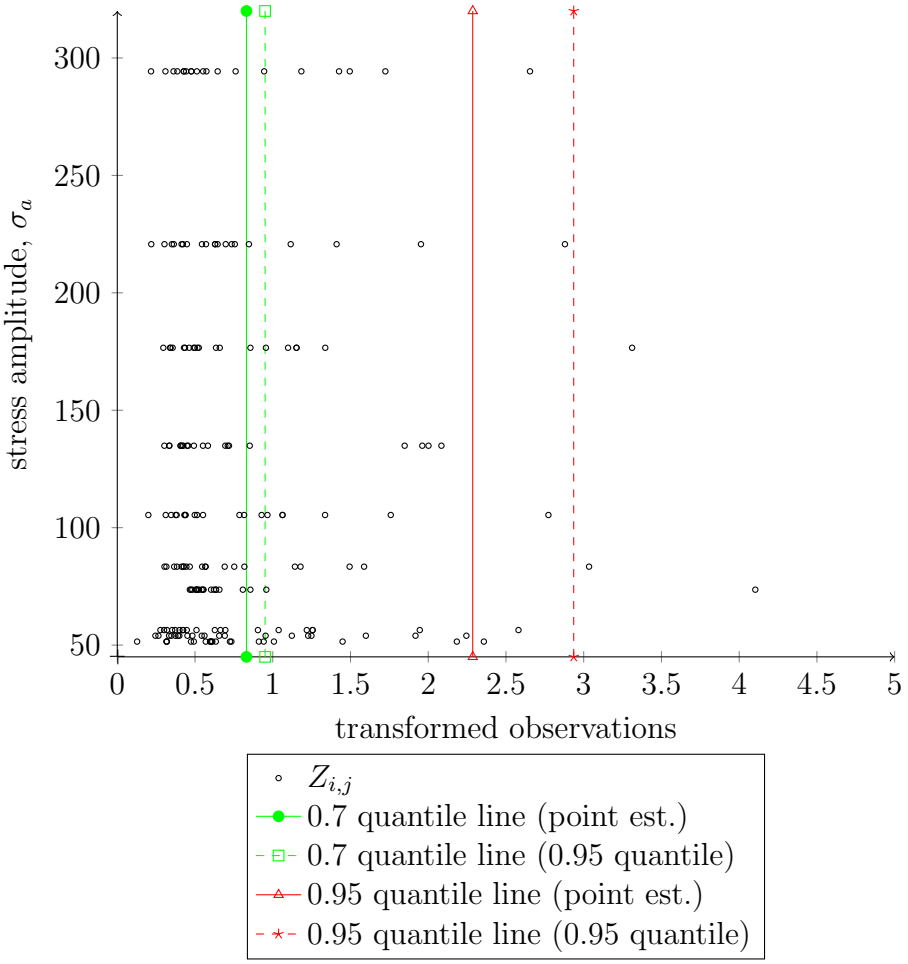


Figure 10: Plot of the transformed observations using Equation (6.1). Additionally the point estimators of the 0.7 and 0.95 quantiles as well as their 0.95 quantile values are displayed.

Table 25: Transformed observations - Pairwise Kolmogorov-Smirnov test

stress	294.3	220.7	176.6	134.9	105.4	83.4	73.6	56.4	54	51.5
294.3	1.00	0.98	0.98	0.98	0.98	0.98	0.17	0.82	0.98	0.56
220.7	0.98	1.00	0.82	0.98	0.82	0.82	0.17	0.98	0.98	0.98
176.6	0.98	0.82	1.00	0.98	1.00	0.98	0.17	0.82	0.98	0.56
134.9	0.98	0.98	0.98	1.00	0.82	0.98	0.03	0.98	0.82	0.56
105.4	0.98	0.82	1.00	0.82	1.00	0.98	0.08	0.98	0.98	0.56
83.4	0.98	0.82	0.98	0.98	0.98	1.00	0.03	0.98	0.98	0.34
73.6	0.17	0.17	0.17	0.03	0.08	0.03	1.00	0.03	0.08	0.56
56.4	0.82	0.98	0.82	0.98	0.98	0.98	0.03	1.00	1.00	0.56
54	0.98	0.98	0.98	0.82	0.98	0.98	0.08	1.00	1.00	0.57
51.5	0.56	0.98	0.56	0.56	0.56	0.34	0.56	0.56	0.57	1.00

Pairwise Kolmogorov-Smirnov test p-values of the transformed observations using Equation (6.1), see Figure 10. This values describe the possibility of measuring the observed data or even extremer data if H_0 (the two datasets are from the same distribution) holds.

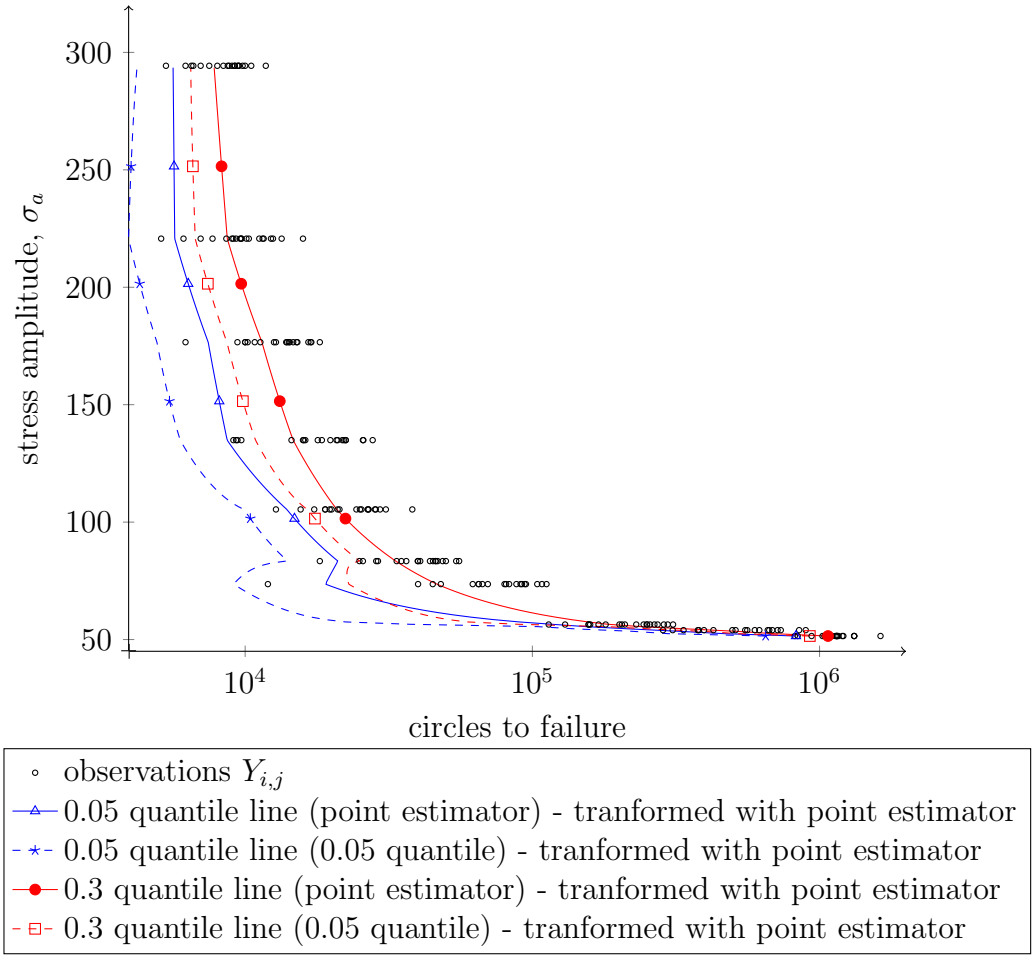


Figure 11: Plot of the observations with quantile lines, gained from re-transforming (with point estimators) the quantiles of Figure 10 and piecewise linear behaviour of ω between the observed values $\omega_i, \quad 1 \leq i \leq 10$.

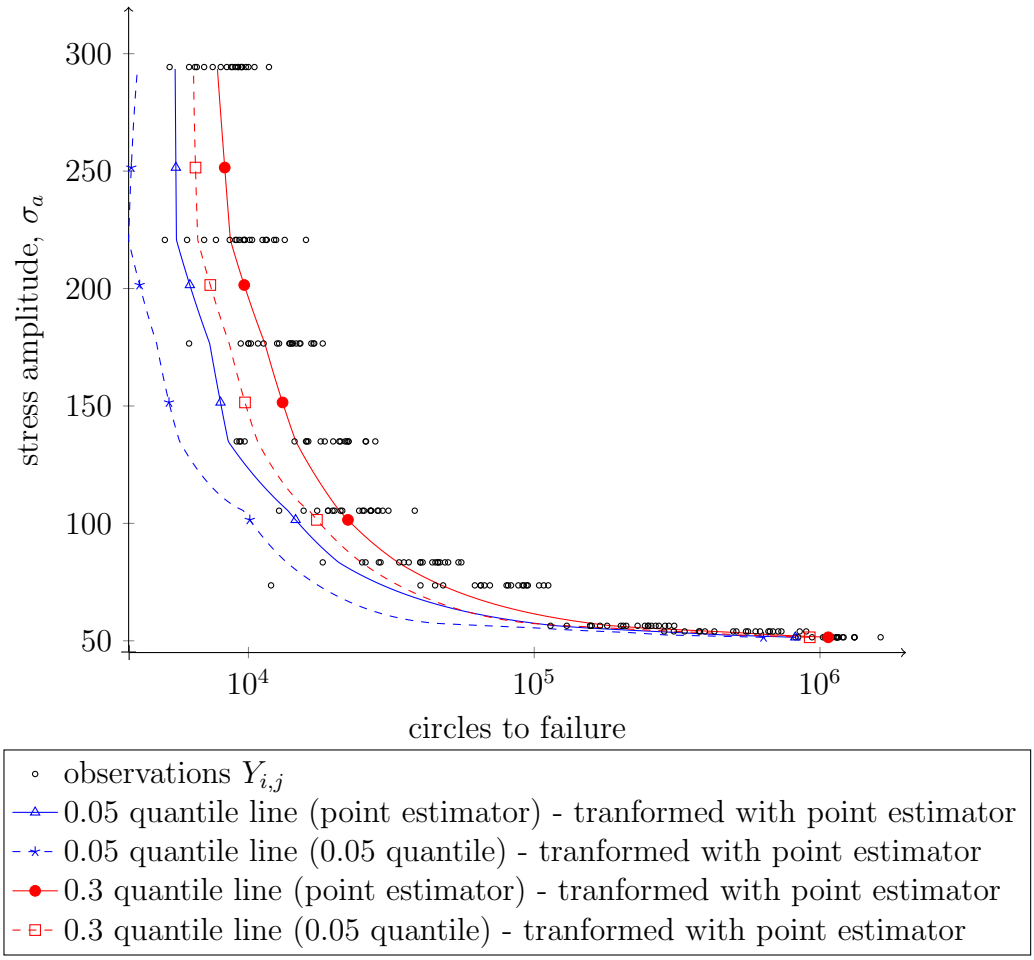


Figure 12: Plot of the observations with quantile lines, gained from re-transforming (with point estimators) the quantiles analog to Figure 11, with stress level 73.6 removed.

7. Sample Application

In this section we will again work with the AAW dataset, taken from Shen (1994, pp. 262); It can be found in Table 1.

When analysing the distribution at each of the ten given stress levels with the Kolmogorov-Smirnov test, discussed in Section 4.4, we obtain the following p-values for the Normal, Gamma, Weibull, and GEV distribution given in Table 26.

Table 26: Kolmogorov-Smirnov test p-values for the AAW data-set

stress	p-values					
	GEV	Gamma	Weibull	Normal distribution		
				y	$\log(y)$	$y^{-1/2}$
294.3	0.909	0.661	0.985	0.833	0.583	0.471
220.7	0.925	0.931	0.936	0.989	0.810	0.603
176.6	0.611	0.585	0.965	0.788	0.544	0.461
134.9	0.676	0.672	0.789	0.841	0.735	0.454
105.4	0.857	0.931	0.956	0.947	0.854	0.713
83.4	0.650	0.690	0.928	0.874	0.567	0.415
73.6	0.610	0.615	0.967	0.895	0.310	0.086
56.4	0.543	0.709	0.962	0.869	0.697	0.636
54.0	0.880	0.923	0.944	0.917	0.946	0.783
51.5	0.645	0.848	0.507	0.717	0.799	0.689

The p-values of the Kolmogorov-Smirnov test are shown for each stress level once for the GEV, Gamma, and Weibull distribution and once for the Normal distribution with three different axis-transformations of Y. The $\log(y)$-transformation correspond to the log-normal distribution, y stands for the identical transformation. These values describe the possibility of measuring the observed data or even extremer data if H_0 holds.

7.1 Analysis of Variance

In this section we ignore the ratio scale properties and treat the stress-level as a factor. This is especially useful to investigate the properties of the data and to check base requirements for possible data models, such as the Linear Model. For example, the Levene test (Nordstokke and Zumbo, 2010) can be used to decide on homogeneous variances for all stress-levels. The values for different transformations of the response is given in Table 27. Only the log-transformation seems to justify a homogeneous variance assumption; A rejection of the null hypothesis (homogeneous variances among all stress-levels) would be a wrong decision in approximately one of ten cases.

7.2 Adaption of the Linear Model

Subsequently, a Linear Model, as introduced in Chapter 5 will be adapted to the data-set given in Table 1. Hence we are investigating the model $Y = X\beta + \epsilon$ with

Table 27: Levene test p-values for the AAW data-set

transformation	p-values
$f(x) = x$	$2 \cdot 10^{-24}$
$f(x) = \log(x)$	0.098
$f(x) = 1/\sqrt{(x)}$	$5 \cdot 10^{-5}$

The p-values of the Levene test are shown for each stress level once for the original axis, the log-transformed axis and, the $1/\sqrt{\cdot}$-transformed axis.

$\epsilon \sim N(0, \sigma^2 I_n)$. This especially means that the data at each stress-level needs to be normal distributed with the same variance. We can use the Kolmogorov-Smirnov test from Section 4.4 to validate the distribution-assumption (see Table 26) and the Levene-test for equal variances (see Table 27) as well as the F-test (see table 28) in the normal-distribution case from Section 4.3 to check the assumption of equal variances.

Table 28: F-test p-values for the AAW data-set: log-transformed data

Y \X	294.3	220.7	176.6	134.9	105.4	83.4	73.6	56.4	54	51.5
294.3	0.50	0.89	0.87	0.99	0.87	0.95	1.00	0.93	0.98	0.15
220.7	0.11	0.50	0.46	0.89	0.45	0.65	1.00	0.60	0.83	0.01
176.6	0.13	0.54	0.50	0.91	0.49	0.69	1.00	0.64	0.86	0.02
134.9	0.01	0.11	0.09	0.50	0.09	0.20	0.92	0.16	0.39	0.00
105.4	0.13	0.55	0.51	0.91	0.50	0.70	1.00	0.65	0.86	0.02
83.4	0.05	0.35	0.31	0.80	0.30	0.50	0.99	0.45	0.72	0.00
73.6	0.00	0.00	0.00	0.08	0.00	0.01	0.50	0.01	0.04	0.00
56.4	0.07	0.40	0.36	0.84	0.35	0.55	0.99	0.50	0.76	0.01
54.0	0.02	0.17	0.14	0.61	0.14	0.28	0.96	0.24	0.50	0.00
51.5	0.85	0.99	0.98	1.00	0.98	1.00	1.00	0.99	1.00	0.50

The p-values from the f-test, see Section 4.3, testing $H_0 : \sigma_X = \sigma_Y$ against $H_1 : \sigma_X > \sigma_Y$ with log-transformed responses Y. This values describe the possibility of measuring the observed data or even extremer data if H_0 holds.

The simplest approach would be taking the stress-level as Y-values and the cycles to failure as x-values. Since a comparison of Table 27 leads to the conclusion, that log-transformed responses Y fit the model variance assumption better we will use the transformed responses. In addition we will also transform the x axis with $f(x) = \frac{1}{x}$ which support the linear expectation assumption of the model much better (see Figure 13). When we now try to identify useful predictors for the model we could take the (transformed) stress w to the power of d for $0 \leq d \leq 9$ into account. This leads us to the model lm_d:

$$Z_i = \log(Y_i) = \sum_{j=0}^{d} \left(\frac{1}{x_i}\right)^j \beta_j + \epsilon_i = \sum_{j=0}^{d} w_i^j \beta_j + \epsilon_i, \qquad 1 \leq i \leq n, \tag{7.1}$$

with $\epsilon_i \sim N(0, \sigma^2)$, $k = 10$ stress level, $m = 20$ iterations per stress level and hence

$n = k \cdot m = 200$ equations. Each pair (Y_i, x_i) correspond to one entry in Table 1.

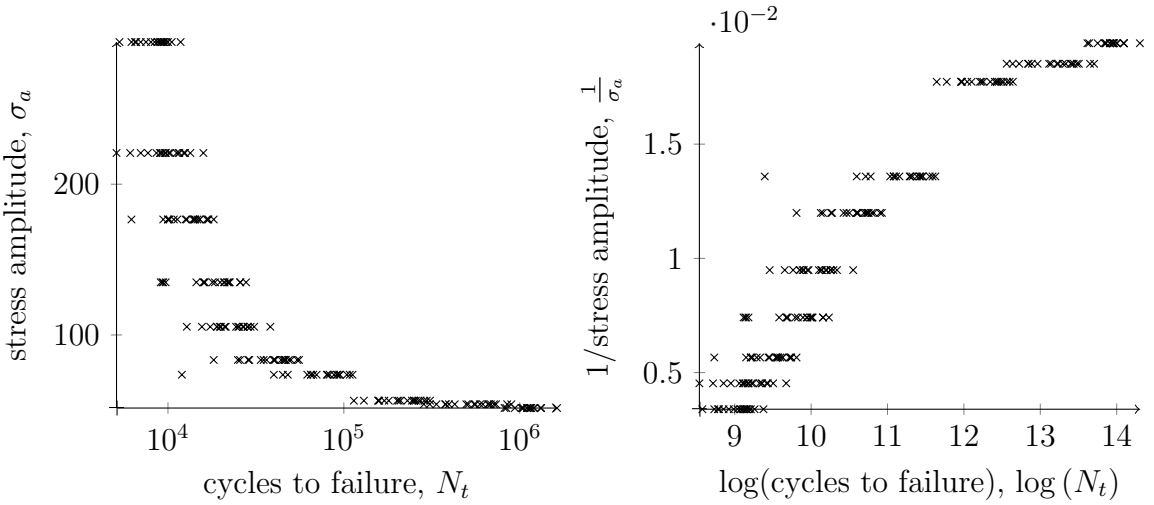


Figure 13: Plots of the AAW Dataset. The left plot shows the original stress-levels, the right one the inverse transformed stress-levels.

Since only ten different stress-levels are observed, a model with an intercept and nine different slope parameters would fit the data perfectly, i.e. at each given stress-level, the predicted response would be the mean of the observed responses at this stress-level. When we look at the other R^2 and R^2_{adj} values in Table 29, corresponding to the model i with the first i powers of x included we can observe, that there is nearly no change in value after model two or three. The corresponding regression lines are shown in Figure 14.

Table 29: $R^2 \setminus R^2_{adj} \setminus$AIC values for the AAW data-set

model	R^2	R^2_{adj}	AIC
lm_0	0.000	0.000	776
lm_1	0.911	0.910	294
lm_2	0.951	0.951	176
lm_3	0.959	0.959	141
lm_4	0.963	0.962	124
lm_5	0.965	0.964	117
lm_6	0.965	0.964	117
lm_7	0.966	0.965	111
lm_8	0.967	0.966	109
lm_9	0.968	0.967	104

The model stated here with lm_i correspond to the model with intercept and i slope parameters - the first i powers of $w = \frac{1}{x}$.

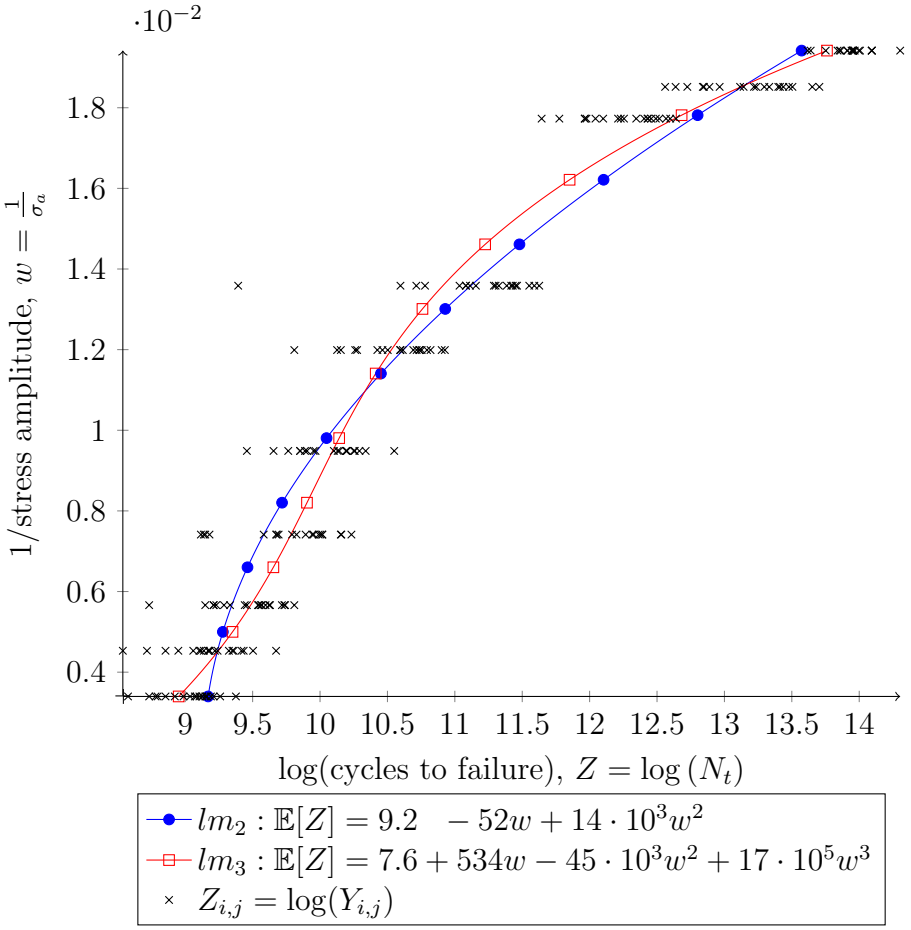


Figure 14: Plot of the transformed AAW Dataset with LM regression line.

Looking at the analysis plots for model lm_2, Figure 15, we can see in the upper left corner the *Residual vs Fitted* plot. Since the variation seems to be equal for all fitted values the constant variance assumption seems to be met. A divergence of the fitted-value structure can be observed - at some stress-levels, the residuals do not vary around zero. On the *QQ-plot* of the standardized residuals at the upper right corner one can see differences to the normal distribution quantiles at the left and right of the plot. The values in the center of the plot seem to lie at the theoretical quantile line. Since the values at the right and left are under the theoretical quantile line we have a negative skewness, the mass of the distribution is focused on the right. The *Scale-Location* plot in the lower left corner shows the fitted values against the square of the standardized residuals. Their distribution should be close to a Normal distribution; The Saphiro-Wilk test (Shapiro and Wilk, 1965) for normality results in a p-value of $5.4 \cdot 10^{-6}$. There is no indication against the homogeneity of variance. In the lower right corner, the *Residual vs Leverage* plot shows high influential data-points. In our case, there are no such points, mainly because we discuss a dataset, where we have multiple data-points with the same value for the predictors.

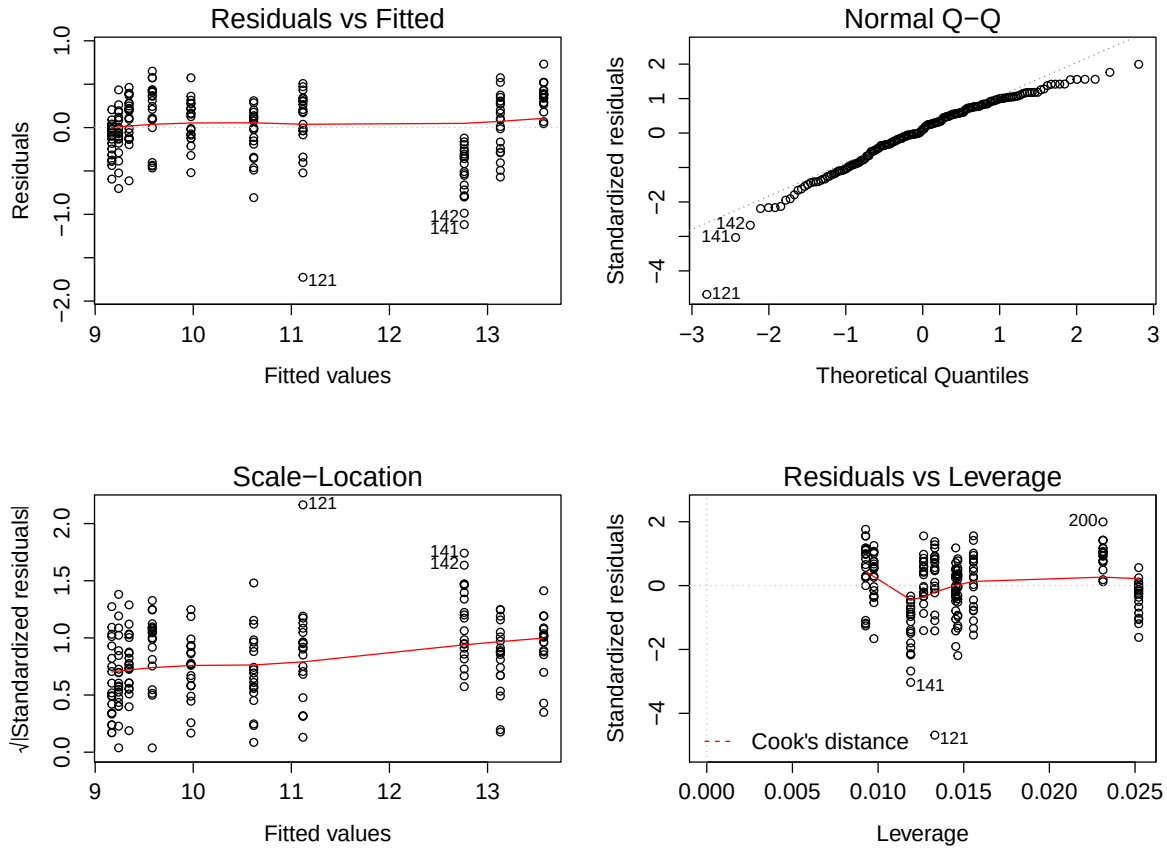


Figure 15: Plot of the residuals, QQ-plot and Cook's distance for the AAW data-set with w and w^2 as predictors.

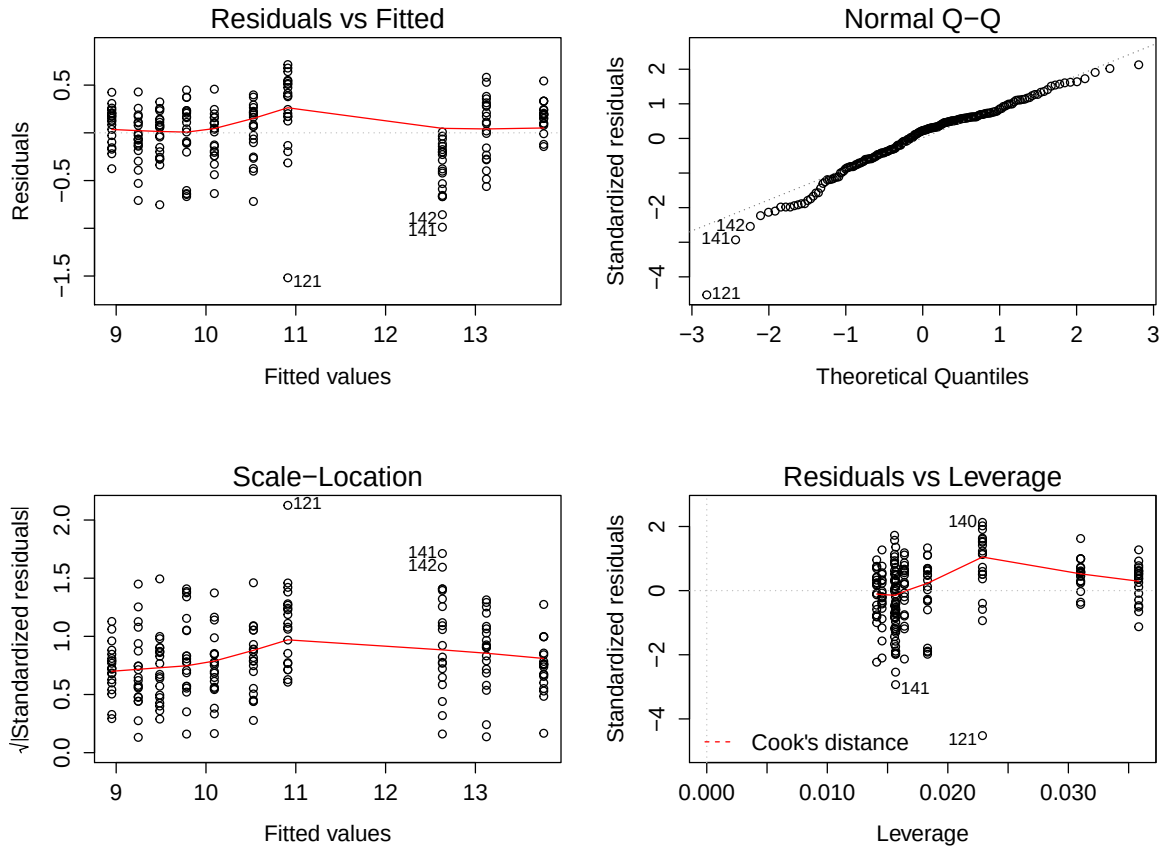


Figure 16: Plot of the residuals, QQ-plot and Cook's distance for the AAW data-set with w, w^2 and w^3 as predictors.

The analysis plot for model lm_3, shown in Figure 16, shows a very similar structure. The only thing which seems to be improved is the *Residual vs Fitted* plot. The Saphiro-Wilk test (Shapiro and Wilk, 1965) for normality results in a p-value of $1.4 \cdot 10^{-5}$. Except for the third highest stress-level, the predicted values fit the observations good. Due to that, also the *Scale-Location* plot show slight improvements in contrast to this plot when using model lm_2.

The confidence and prediction intervals, according to Section 5, are shown in Figure 17 and Figure 18 for the original data. A confidence level $1 - \alpha = 0.95$ for both intervals and both plots was chosen. We can say if the Linear Model assumptions hold, that there is a chance of 95% the true population parameter is contained in the confidence interval. That means, if we would repeat the procedure of gathering data and calculating the confidence interval for another 99 times there would the true population parameter be in 95 of the 100 confidence intervals on average. The prediction interval is an estimation of a range which contains a new observation with 95% certainty.

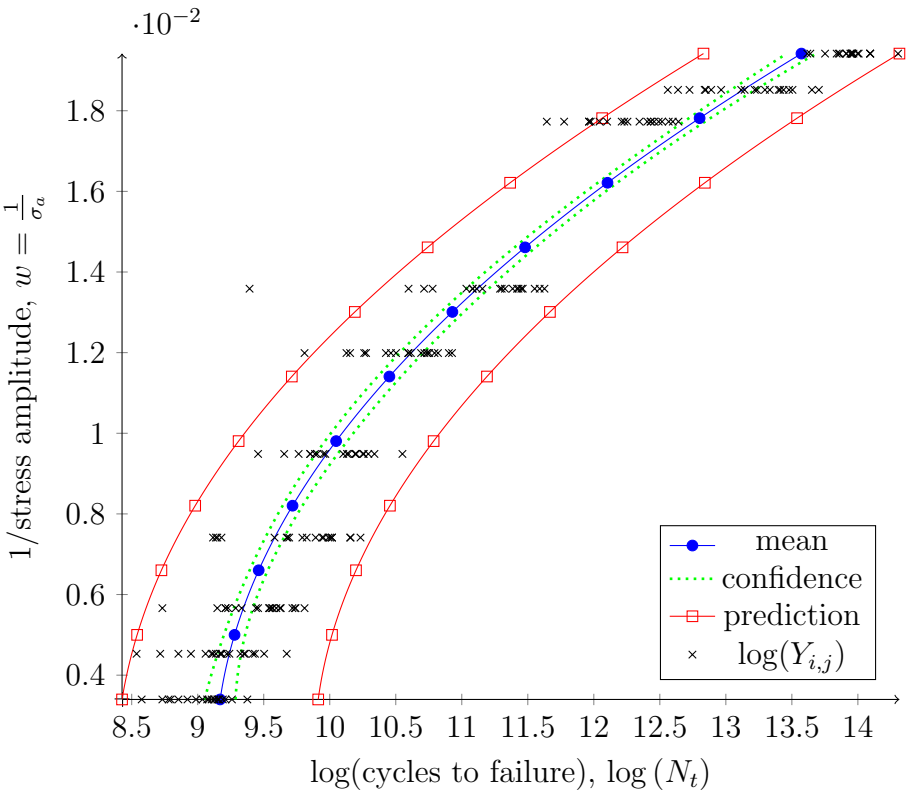


Figure 17: Plot of the LM confidence and prediction intervals for the AAW data-set with w and w^2 as predictors and $1 - \alpha = 0.95$. The model response (mean) is also shown.

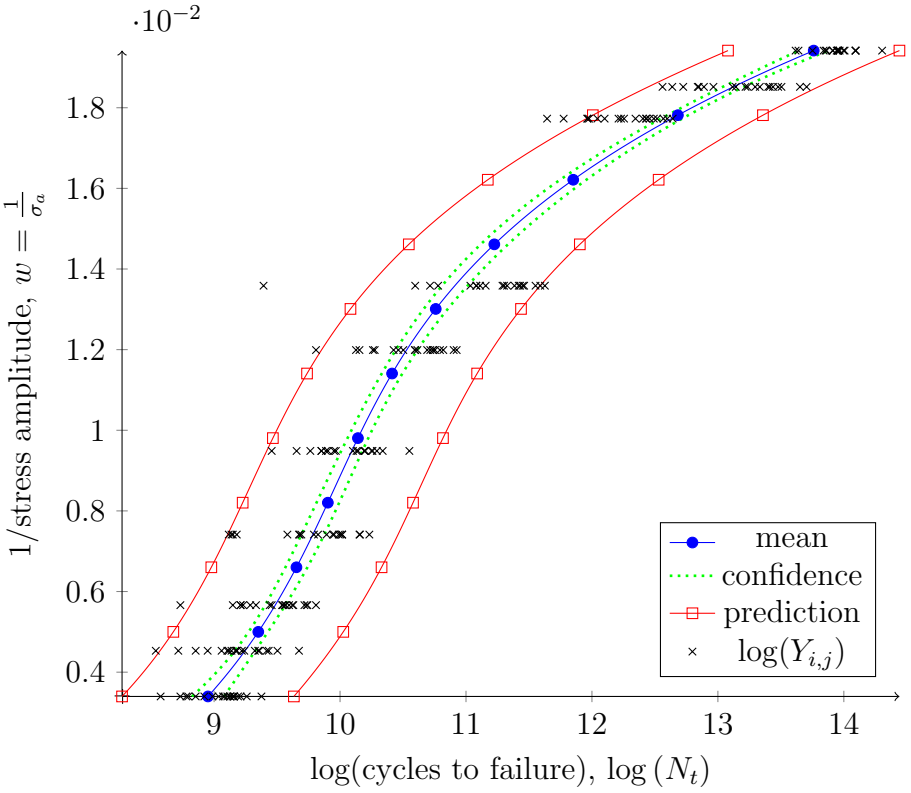


Figure 18: Plot of the LM confidence and prediction intervals for the AAW data-set with w, w^2 and w^3 as predictors and $1 - \alpha = 0.95$. The model response (mean) is also shown.

Table 30: p-values for the AAW data-set - m_3

stress	mean	sigma^2	pval
294.3	8.951	0.115	0.169
220.7	9.246	0.115	0.463
176.6	9.485	0.115	0.566
134.9	9.784	0.115	0.668
105.4	10.093	0.115	0.499
83.4	10.528	0.115	0.503
73.6	10.911	0.115	0.001
56.4	12.632	0.115	0.000
54.0	13.123	0.115	0.612
51.5	13.761	0.115	0.000

Resulting parameter estimates with KS-test p-values for the model containing three predictors.

7.3 Generalized Linear Model Approach

We have seen, that the assumption of normally distributed responses is not satisfied and the Gamma-distribution could be a suitable choice for the untransformed responses, as shown in Table 26. Since the Gamma distribution is member of the Exponential family, see Example 3.2, we can apply a GLM model with Gamma-distributed responses. We are now working with the model glm_d:

- $Y_i \sim Gamma(k, \lambda_i)$,
- $\eta_i = X_i^T \beta = \sum_{j=0}^{d} x_i^j$,
- $g(\mu_i) = \eta_i$.

In the context of an exponential family $Exp(\theta_i, \phi)$, whereas the random variables Y feature the density function

$$f(y_i, \theta_i, \phi) = e^{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + h(y_i, \phi)},$$

we know from Example (3.2), that this density can be linked to the density function of a $Gamma(k, \lambda_i)$ distributed random variable with the equations $a(\phi) = \phi = \frac{1}{k}$, $\theta_i = -\frac{\lambda_i}{k}$, $b(\theta_i) = -\log(-\theta_i)$ and $h(y_i, \phi) = \frac{1}{\phi}\log(\frac{1}{\phi}) + (\frac{1}{\phi} - 1)\log(y_i) - \log(\Gamma(\frac{1}{\phi}))$.

We now need to decide with which link function $g()$ we should work. Since $g(\mu_i)$ is modelled with a linear function of the predictors we can use the sample mean of each stress level to decide on $g()$. Since $b(x) = -\log(-x)$ holds we observe $b'(x) = -\frac{1}{x}$ and get the canonical link as $g(x) = (b'(x))^{-1} = -\frac{1}{x}$. This leads us to the model glm_d:

$$Y_i \sim Gamma(k, \lambda_i), \quad \mathbb{E}[Y_i] = \mu_i = -\frac{1}{\sum_{j=0}^{d} x_i^j}. \tag{7.2}$$

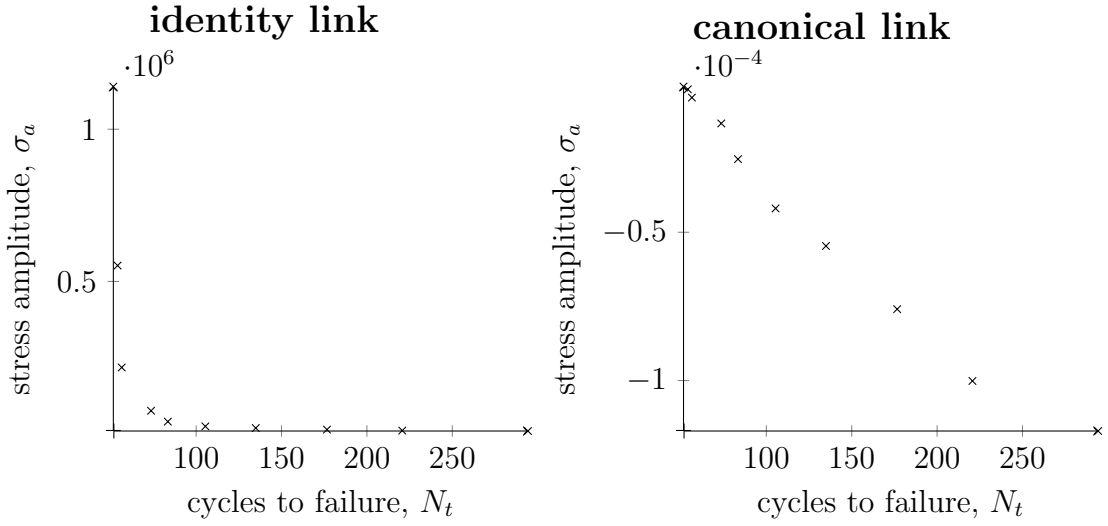


Figure 19: Plot of the identity link $g(x) = x$ and the canonical link $g(x) = -\frac{1}{x}$ evaluated at the observed sample means.

As shown in Figure 19 the canonical link seems to be suitable to map the sample mean to a linear function. We can now calculate $\hat{\beta}$, the estimate for β with the IRLS-procedure and use it to get to the estimates for the parameter θ_i with $\hat{\theta}_i = g(b'(\hat{\theta}_i)) = x_i^T\hat{\beta}$. Through Definition 3.5 we obtain an estimate for the scale parameter ϕ:

$$\hat{\phi} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{V(\hat{\mu}_i)} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{b''((b'(\hat{\mu}_i))^{-1})} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{\hat{\mu}_i^2},$$

which also leads us to an estimate for the distribution parameter $\hat{k} = \frac{1}{\hat{\phi}}$. In combination with the identity $\theta_i = -\frac{\lambda_i}{k}$ we get the second distribution parameter estimate with $\hat{\lambda}_i = -\hat{\theta}_i\hat{k}$. The p-values for these estimates are shown in Table 31 for different numbers of predictors.

Table 31: Kolmogorov-Smirnov test p-values - GLM model

stress	glm_1	glm_2	glm_3	glm_4	glm_5	glm_6	glm_7	glm_8
294.3	0.00	0.14	0.47	0.54	0.64	0.64	0.64	0.64
220.7	0.36	0.83	0.44	0.70	0.74	0.70	0.74	0.76
176.6	0.56	0.81	0.15	0.19	0.76	0.85	0.71	0.66
134.9	0.62	0.64	0.52	0.18	0.38	0.33	0.62	0.60
105.4	0.01	0.05	0.10	0.44	0.68	0.73	0.97	0.48
83.4	0.01	0.12	0.08	0.20	0.58	0.54	0.63	0.13
73.6	0.15	0.02	0.05	0.04	0.01	0.01	0.03	0.01
56.4	0.13	0.48	0.18	0.04	0.01	0.01	0.01	0.07
54.0	0.00	0.00	0.00	0.01	0.03	0.03	0.02	0.09
51.5	0.04	0.01	0.03	0.08	0.13	0.13	0.11	0.15

The p-values from the Kolmogorov-Smirnov test when using parameter estimators gained from the GLM model. The model $m_i \quad i \in \{1, ..., 8\}$ uses the first i powers of the stress level and the intercept for estimating η. This values describe the possibility of measuring the observed data or even extremer data if H_0 holds. All 200 observations are used for this fit.

Comparing the models with different numbers of predictors according to the AIC-criteria as discussed in Section 3.2.2 leads to the model glm_5, see Table 32. When fitting the model to the observations the AIC recommends the model with an intercept and the first five powers of the stress level for the linear predictor η.

Table 32: AIC values for different number of predictors

model	glm_1	glm_2	glm_3	glm_4	glm_5	glm_6	glm_7	glm_8
AIC	4468.85	4459.07	4453.74	4450.59	**4448.38**	4450.30	4451.17	4449.25

AIC values for different GLMs using the Gamma distribution. The model $glm_i \quad i \in \{1, ..., 8\}$ uses the first i powers of the stress level and the intercept in the model.

In Table 33 we can see the parameter estimators for the gamma distribution for each stress level. Once the parameters are estimated for each stress level without any

restriction using the method of moments. The other estimation series uses the GLM recommended by the AIC criterion. The related quantile lines which are based on the gained point estimators of the generalized linear model glm_5 can be found in Figure 21. When looking at the area between the 0.025 and 0.975 quantile lines one can see that exactly 5% (10 observations out of 200) are outside. The same information with inverted stress scale can be found in Figure 21; It can be used to compare it to the Linear Models. The corresponding residual plot can be found in Figure 20.

Table 33: Parameter estimations - GLM models

	individual estimators			GLM fit - glm_5		
stress	$\hat{k}_{ind}$	$1/\hat{\lambda}_{ind}$	$\hat{\mu}_{ind}$	$\hat{k}_{single}^{glm_5}$	$1/\hat{\lambda}_{single}^{glm_5}$	$\hat{\mu}_{single}^{glm_5}$
294.3	29.44	290.28	8545	11.58	737.97	8543
220.7	16.30	612.75	9985	11.58	862.25	9982
176.6	19.20	685.96	13170	11.58	1147.11	13280
134.9	10.70	1710.65	18305	11.58	1519.38	17590
105.4	16.58	1437.22	23825	11.58	2179.75	25236
83.4	15.46	2551.43	39440	11.58	3713.51	42993
73.6	9.26	8109.05	75100	11.58	5535.52	64087
56.4	15.07	14416.06	217300	11.58	24462.25	283213
54.0	10.23	53993.53	552150	11.58	39998.35	463083
51.5	42.26	26983.65	1140200	11.58	101058.00	1170006

Gamma distribution parameter estimations using the moment estimators for the individual parameter estimations for each stress level and the GLM estimations for all observations are shown in this table. The chosen model for the GLM estimations are the one recommended by the AIC criterion, namely model glm_5. Also, the estimated mean μ is shown.

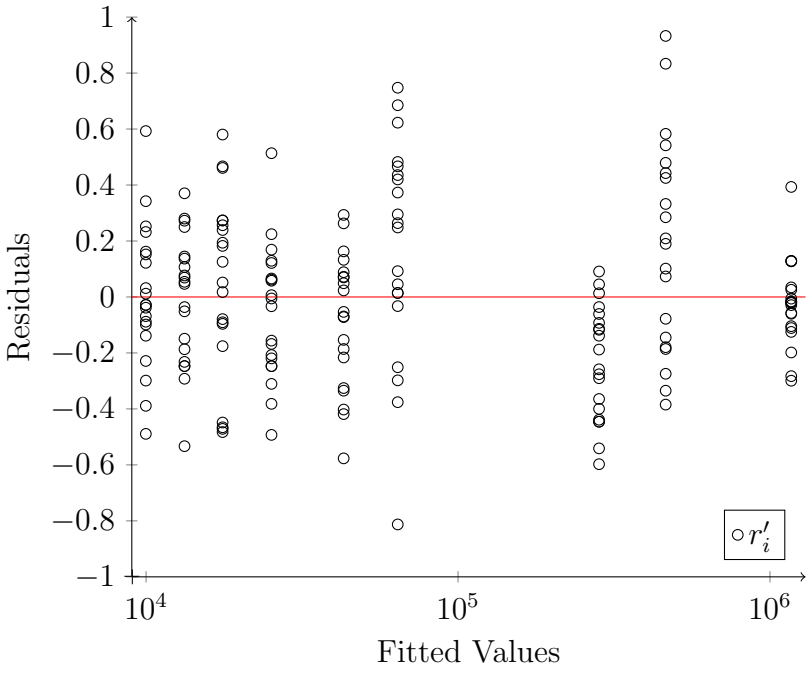


Figure 20: Plot of the fitted values against the scaled residuals ($r'_i = \frac{Y_i - \mu_i}{\mu_i}$) of the GLM model glm_5.

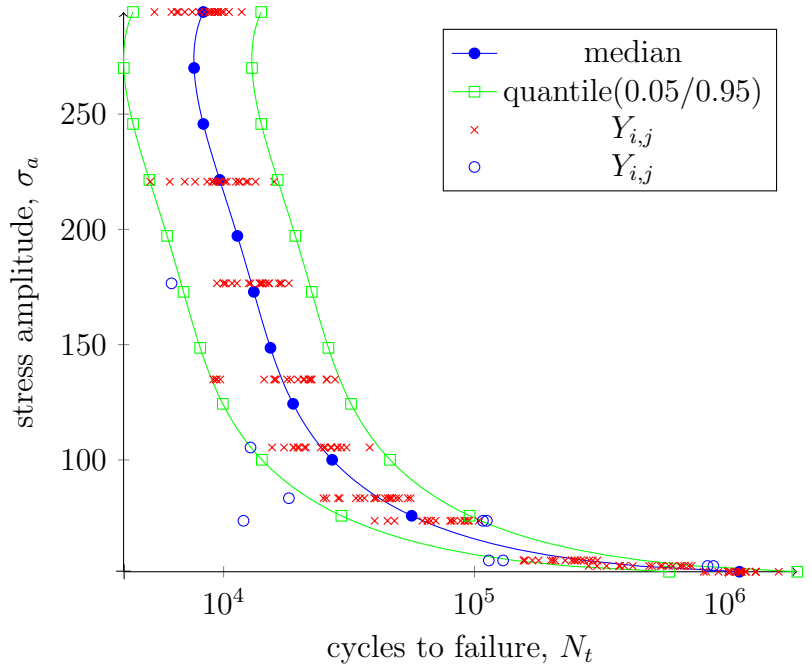


Figure 21: Plot of the 0.05 and 0.95 quantile lines for the GLM m_5 fitting all observations (green, dashed, squares). The points represent the observations within the 90% quantile area of the GLM glm_5 (red, stars) and outside this area (blue, circles).

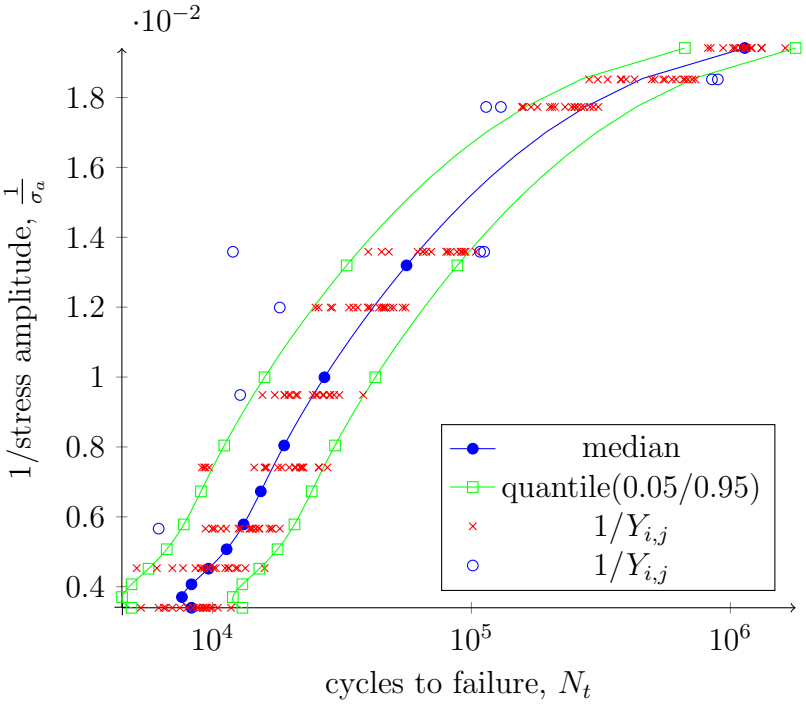


Figure 22: Plot of the 0.05 and 0.95 quantile lines for the GLM m_5 fitting all observations (green, dashed, squares) with inverted stress axis. The points represent the observations within the 90% quantile area of the GLM glm_5 (red, stars) and outside this area (blue, circles).

8. Conclusion

In this thesis we discussed multiple models with different strength and weaknesses. We started with the widely used Linear Model and Generalized Linear Model approaches as representatives of the status quo. Additionally, we investigated a model based on parameter regression adjusted to the special context of fatigue data. As the last discussed model we introduced the standardizing model.

For the Linear Model, we identified the polynomial model lm_2 with two and lm_3 with three predictors as best suitable for modelling the given data. Additionally, we applied transformations to both axes; The stress axis was inversely transformed and the cycle to failure axis was log-transformed. The corresponding model is given by Equation (7.1).

The Generalized Linear Model was fitted to the data using the Gamma distribution and the canonical link function with the linear predictor as a polynomial of degree five, resulting in model glm_5. It was chosen based on the best AIC value. The model is defined by Equation (7.2).

The adaption of parameter regression to the context of fatigue data lead us to the model pr. In this context the distribution type is fixed, only the parameters of the distribution vary based on the stress level. We introduced a set of five parameters, called structure parameters which can be used in combination with the stress level to express the distributional parameters. By applying the maximum likelihood method we gained the structure parameters, which define the GEV distribution parameters for each stress level. This model is given in Equation (5.1).

Based on the idea of standardizing we gain the model sa. When using this approach we assume the observations to be identically and independently distributed after a transformation; This transformation may depend on a transformation-quantity calculated by all observations of this stress level, as in our case. Then we can estimate the parameters and re-transform the quantiles gained. In between the stress levels, we can use a piecewise linear function or splines to obtain the transformation-quantity needed for re-transformation. The model is based on Equation (6.2).

We examine these five proposed models based on their ability to represent the AAW dataset. The model abilities will be measured using two different perspectives, called median fit quality and scatter fit quality. The median fit quality describes how good the model can mirror the observed median; The scatter fit quality describes the model's ability to reflect the variation of the data. We want to clarify once again, that we are not making general statements, but only statements for this very special situation.

8.1 Median Fit

We transformed the axis of all models to be able to compare them visually. A suitable axis transformation seems to be the inverse transformation for the stress level; This allows an easy distinction of the tight arranged low-stress observations. For the cycles to failure axis, we used a logarithmic scale; This neutralizes the intense varying scatter behaviour. The estimates of the median are given in Figure 23. A remarkable observation is, that the GLM glm_5 is the only model not having a monotone median curve. This can be observed in Figure 23 for low y-axis values ($\frac{1}{\sigma_a} \approx 0.4 \cdot 10^{-2}$).

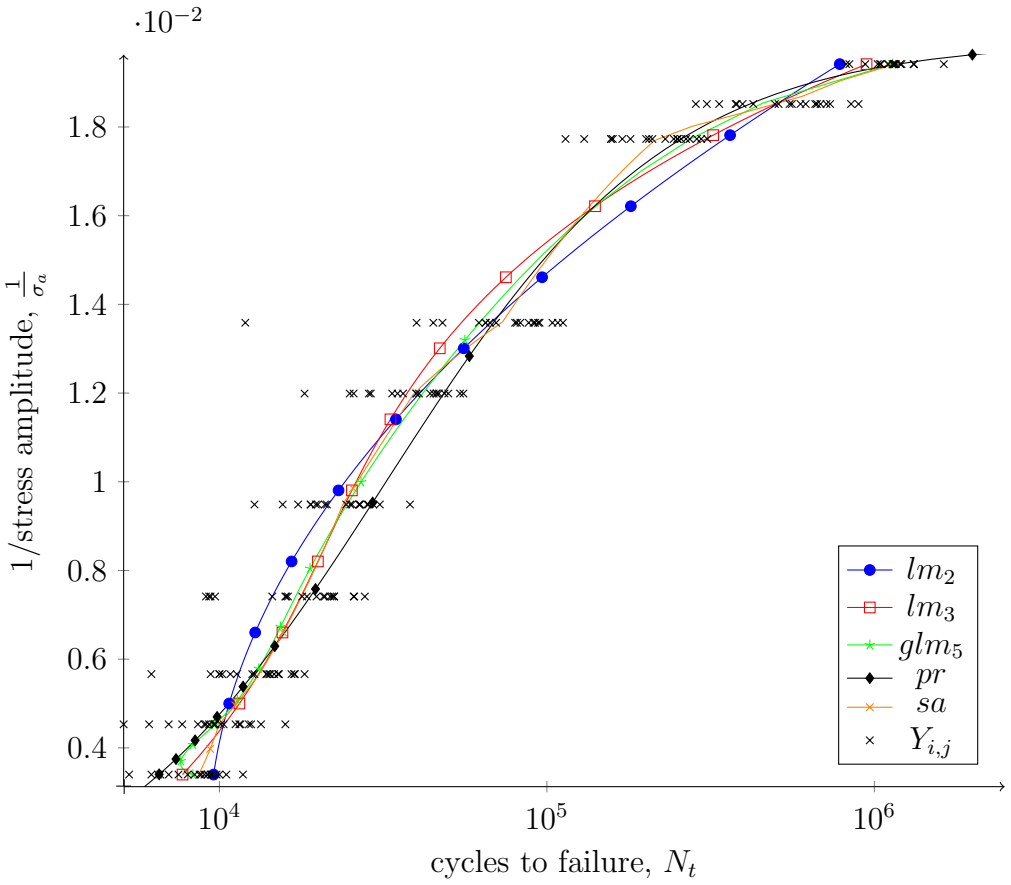


Figure 23: Plot of the transformed AAW Dataset with median regression line of the Linear Models lm_2 and lm_3 (see Equation (7.1)), the Generalized Linear Model glm_5 (see Equation (7.2)), the parameter regression model pr (see Equation (5.1)), and the standardizing approach model sa from Section 6.5.

The scaled deviation between the model median and the observations allow a much better analysis. Therefore, we have plotted the value of $\frac{Y_{i,j}-med_i}{med_i}$ (called residuals in this context) against the model prediction of the median med_i (called fitted values), see Figure 24. For scaling purposes we divided the deviation by the fitted value; This

results in visually similar spread residuals.

The Linear Models lm_2 and lm_3 are fitting the data appropriate, except for two stress levels. The stress level associated with the third highest fitted values shows only negative residuals; The stress level associated with the highest fitted values shows nearly just positive residuals. This is a signal, that the median, which should feature approximately as much negative residuals as positive residuals, is not predicted appropriate. We can say, that the residuals for model lm_3 are much more uniformly spread around zero.

For the Generalized Linear Model glm_5 we observe a better mean fit quality as for the Linear models. The stress levels associated with the third highest and the highest fitted values are much better represented with the model glm_5. However, the stress levels associated with the third highest fitted value is still not adequate.

The Parameter Regression Model *pr* seems to reproduce the median curve in a similar quality as the Linear Models. Although there are multiple stress levels, for which the residuals seem to have mostly the same sign, there is no stress level, for which we have only positive or negative residuals. The unequally spread residuals are not necessarily disturbing, due to the changing distribution over the stress level for this model.

The Standardizing Approach leading to the model *sa* features an excellent mean fitting quality due to the model characteristics. This is always the case if we have equally distributed observations over all stress levels after the transformation.

Due to this observations, we are ranking the Standardizing Approach Model *sa* first according to the median fit quality. The Generalized Linear Model glm_5 and the Parameter Regression Model *pr* seem to fit the median of the observed data similarly and better as the two Linear Models lm_2 and lm_3. Between the two Linear Models, we can clearly decide in favour of the model lm_3 including an additional predictor in comparison to the model lm_2.

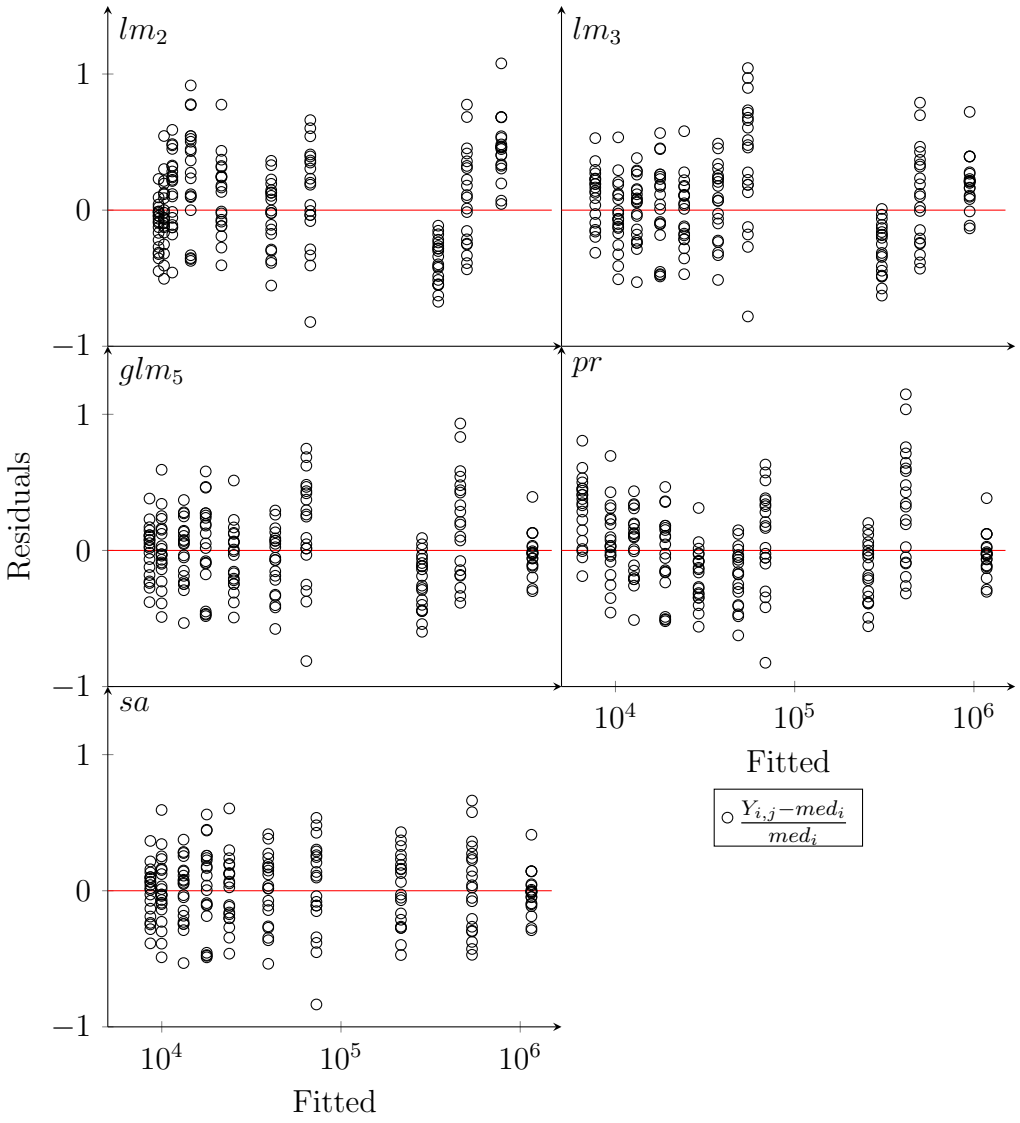


Figure 24: Plot of the transformed AAW Dataset; Point estimates for the Linear Models lm_2 and lm_3 (see Equation (7.1)), the Generalized Linear Model glm_5 (see Equation (7.2)), the parameter regression model pr (see Equation (5.1)), and the standardizing approach model sa from Section 6.5 are used to calculate the Residuals. In this context we used the median to centralize and scale the observations.

8.2 Scatter Fit

To compare the variation of the different models we examine the quantile lines, given in Figure 25. We can see a constant difference between the quantile lines for the Linear Models on the logarithmic scale. The behaviour of the quantile lines for the Parameter Regression Model and the GLM is similar to the behaviour of the lm_3 model's quantile line behaviour. The Standardizing Approach Model *sa* is estimating unrealistic quantile lines for one of the stress levels.

To determine the scatter fit quality of the models we examine the number of observations falling within the model quantile regions. We have decided to chose the quantile regions $q_1 = (0, q_{0.25}]$, $q_2 = (q_{0.25}, q_{0.5}]$, $q_3 = (q_{0.5}, q_{0.75}]$, and $q_4 = (q_{0.75}, \infty)$. For each of the regions, we would expect one-quarter of the observations falling into the region. This is true for the observations of each stress level and for all observations. The numbers of observations really falling into the regions are given in Table 34.

If we assume the model quantiles to be the true quantiles, we gain a Multinomial distribution for the observations falling in one quantile region. Let N_i be the random variable for elements falling in the quantile region q_i and n_i the realization of this random variable. Then we can give the probability as

$$\mathbb{P}[N = n] = \mathbb{P}[N_1 = n_1, \cdots, N_4 = n_4] = \frac{n!}{\prod_{i=1}^{4} n_i!} \left(\frac{1}{4}\right)^n, \tag{8.1}$$

for $n = \sum_{i=1}^{4} n_i$ observations. The according probabilities for each individual stress level and each model are listed in Table 34. Additionally the product of those probabilities for each model is listed in the sum/product row of the table. This value represent the probability of gaining exactly these observations for all stress levels, if the corresponding model is true, i.e. the quantile lines are the correct ones.

When comparing the models concerning their scatter fit ability, the Standardizing Approach model fits the data best. Due to the probabilities gained by Equation (8.1), we get the following ranking for the models based on their scatter fit quality: *sa* ($1.2{\cdot}10^{-27}$), glm_5 ($1.2{\cdot}10^{-33}$), lm_3 ($6.6{\cdot}10^{-35}$), *pr* ($1.4{\cdot}10^{-36}$), and lm_2 ($3.3{\cdot}10^{-42}$). As a comparison we can take the maximum probability in this situation, i.e. the probability, when we observe exactly 5 samples falling in each category for every stress level, which is $1.9 \cdot 10^{-20}$.

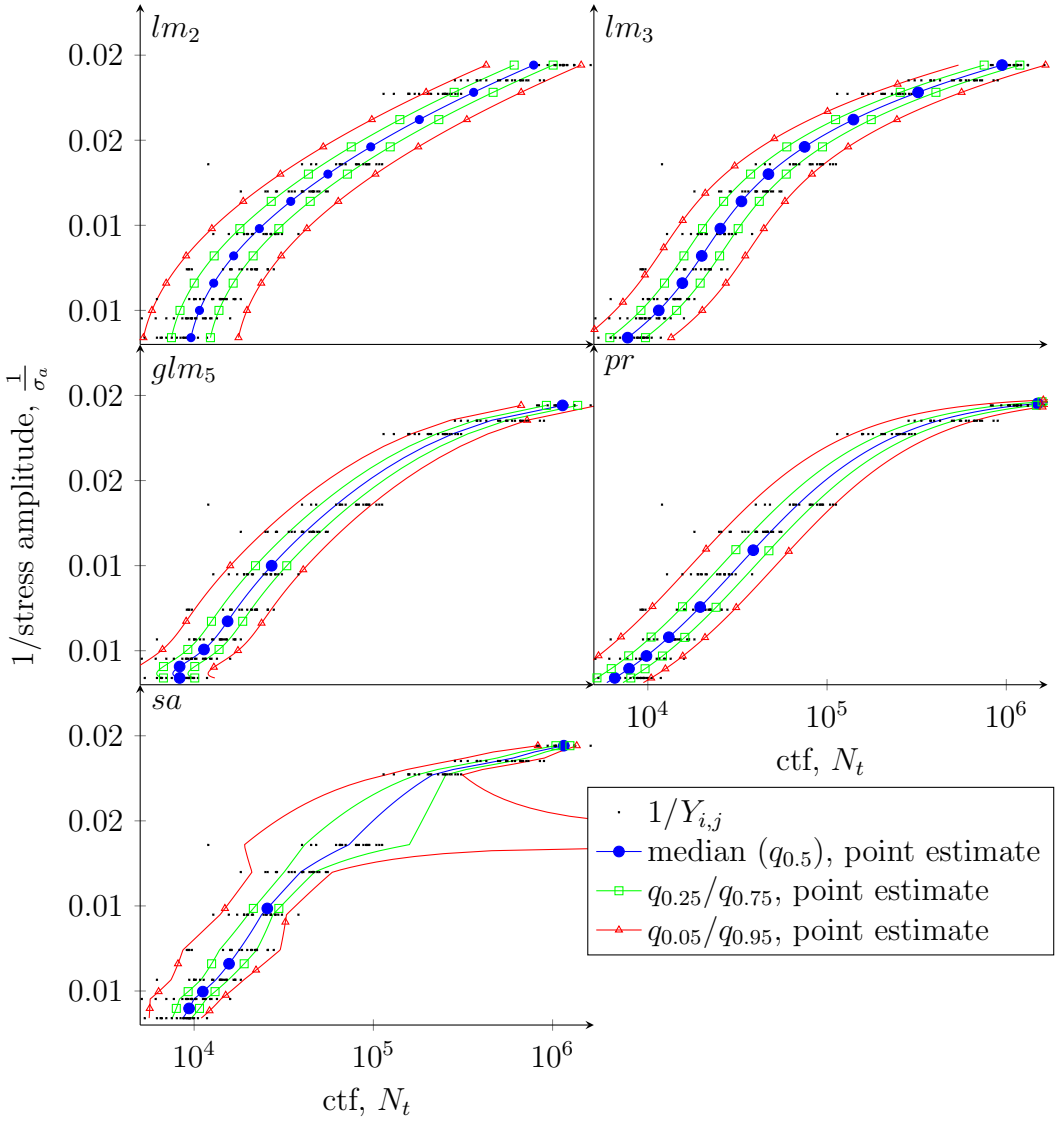


Figure 25: Plot of the transformed AAW Dataset; Point estimates for the Linear Models lm_2 and lm_3 (see Equation (7.1)), the Generalized Linear Model glm_5 (see Equation (7.2)), the parameter regression model pr (see Equation (5.1)), and the standardizing approach model sa from Section 6.5.

8.3 Overall Conclusion

Due to the superior performance of the Standardizing Model *sa* in both categories, the median fit and the scatter fit quality, it seems to be best suitable to describe the AAW data set. The disadvantage of this model is, that we always need multiple observations per stress level. Also the behaviour of the quantile lines (see Figure 25) is questionable.

The Generalized Linear Model glm_5 seems to be the next best model. The non - monotone behaviour of the model's median line is suspect. Although not visible at the plot, the values observed at the stress levels are monotone. This non - monotone behaviour in between the stress levels seems to make this model unsuitable for interpolation in this range of the curve.

The next best models are the Linear Model lm_3 and the Parameter Regression Model pr. It is not possible to prefer one of these models against each other. The Linear Model reproduce the scattering behaviour slightly better than the Parameter Regression Model. Otherwise, the median fit quality is marginally better with the Parameter Regression Model.

The Linear Model lm_2 performs worse compared to the other applied models. This is true for the median fit quality as well as for the scatter fit quality of the model.

Table 34: Observations within quantile areas for all models

Model	Stress	Number of Observations $(0, q_{0.25}]$	$(q_{0.25}, q_{0.5}]$	$(q_{0.5}, q_{0.75}]$	$(q_{0.75}, \infty)$	Probability
	294	5	11	4	0	$1.9 \cdot 10^{-}5$
	221	4	8	6	2	$1.5 \cdot 10^{-}3$
	177	1	6	7	6	$8.4 \cdot 10^{-}4$
	135	4	1	5	10	$2.1 \cdot 10^{-}4$
	105	2	7	6	5	$2.5 \cdot 10^{-}3$
lm_2	83	5	6	7	2	$2.5 \cdot 10^{-}3$
	74	4	4	4	8	$3.9 \cdot 10^{-}3$
	56	16	4	0	0	$4.4 \cdot 10^{-}9$
	54	5	3	5	7	$5.1 \cdot 10^{-}3$
	52	0	0	3	17	$1.0 \cdot 10^{-}9$
	$\sum / \prod$	46	50	47	57	$3.3 \cdot 10^{-}42$
	294	1	5	10	4	$2.1 \cdot 10^{-}4$
	221	4	9	5	2	$1.1 \cdot 10^{-}3$
	177	5	4	7	4	$6.4 \cdot 10^{-}3$
	135	4	4	7	5	$6.4 \cdot 10^{-}3$
	105	5	4	9	2	$1.1 \cdot 10^{-}3$
lm_3	83	5	3	8	4	$3.2 \cdot 10^{-}3$
	74	2	2	4	12	$4.8 \cdot 10^{-}5$
	56	11	8	1	0	$1.4 \cdot 10^{-}6$
	54	6	2	5	7	$2.5 \cdot 10^{-}3$
	52	0	3	11	6	$1.3 \cdot 10^{-}5$
	$\sum / \prod$	43	44	67	46	$6.6 \cdot 10^{-}35$
	294	4	3	11	2	$1.9 \cdot 10^{-}4$
	221	4	5	7	4	$6.4 \cdot 10^{-}3$
	177	5	4	7	4	$6.4 \cdot 10^{-}3$
	135	4	4	3	9	$1.8 \cdot 10^{-}3$
	105	6	4	8	2	$1.6 \cdot 10^{-}3$
glm_5	83	6	5	7	2	$2.5 \cdot 10^{-}3$
	74	4	1	4	11	$9.6 \cdot 10^{-}5$
	56	10	7	3	0	$2.0 \cdot 10^{-}5$
	54	3	4	2	11	$1.9 \cdot 10^{-}4$
	52	2	7	10	1	$6.0 \cdot 10^{-}5$
	$\sum / \prod$	48	44	62	46	$1.2 \cdot 10^{-}33$
	294	0	3	4	13	$2.5 \cdot 10^{-}6$
	221	3	5	7	5	$5.1 \cdot 10^{-}3$
	177	4	4	8	4	$4.0 \cdot 10^{-}3$
	135	5	5	7	3	$5.1 \cdot 10^{-}3$
	105	9	8	2	1	$7.6 \cdot 10^{-}5$
pr	83	8	8	4	0	$5.7 \cdot 10^{-}5$
	74	4	4	4	8	$4.0 \cdot 10^{-}3$
	56	8	7	5	0	$9.1 \cdot 10^{-}5$
	54	2	4	3	11	$1.9 \cdot 10^{-}4$
	52	2	12	5	1	$1.9 \cdot 10^{-}5$
	$\sum / \prod$	45	60	49	46	$1.4 \cdot 10^{-}36$
	294	6	2	8	4	$1.6 \cdot 10^{-}3$
	221	4	7	5	4	$6.4 \cdot 10^{-}3$
	177	6	3	7	4	$4.2 \cdot 10^{-}3$
	135	4	4	7	5	$6.4 \cdot 10^{-}3$
	105	6	3	6	5	$5.9 \cdot 10^{-}3$
sa	83	5	3	8	4	$3.2 \cdot 10^{-}3$
	74	2	7	11	0	$5.5 \cdot 10^{-}6$
	56	6	4	4	6	$7.4 \cdot 10^{-}3$
	54	7	2	4	7	$1.8 \cdot 10^{-}3$
	52	5	8	3	4	$3.2 \cdot 10^{-}3$
	$\sum / \prod$	51	43	63	43	$1.2 \cdot 10^{-}27$

This table is showing how many observations are falling within the quantile areas $(0, q_{0.25}]$, $(q_{0.25}, q_{0.5}]$, $(q_{0.5}, q_{0.75}]$, and $(q_{0.75}, \infty)$ for the models lm_2, lm_3, glm_5, pr, and sa. In the last column the probabilities due to the Binomial distribution, see equation 8.1 are listed for each stress levels individually; The last probability entry for each model is the product of the probabilities for the model's individual stress levels.

A. Mathematical base

A.1 Distributions

Definition A.1. ***(Normal distribution)***
The random variable X follows a Normal distribution $N(0,1)$ if its density is given as:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}. \tag{A.1}$$

In this case the random variable $Y = \sigma \cdot X + \mu$ follows a $N(\mu, \sigma^2)$ normal distribution. Note, that $f_X(x) = f_X(-x)$ holds, i.e. the distribution is symmetrical.

Example A.2. ***(Density of the $N(\mu, \sigma)$ distribution)***
Let $Y \sim N(0,1)$ be standard normal distributed. For $X = \sigma \cdot X + \mu \sim N(\mu, \sigma^2)$ with $\mu \in \mathbb{R}, \sigma \in \mathbb{R}_{>0}$ we can write the cumulative density function as:

$$\begin{aligned} F_X(a) &= \mathbb{P}[X \leq a] = \mathbb{P}\left[\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right] \\ &= \mathbb{P}\left[Y \leq \frac{a-\mu}{\sigma}\right] = F_Y\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

This leads us to:

$$\begin{aligned} f_X(a) &= \frac{d}{da} F_X(a) = \frac{d}{da} F_Y\left(\frac{a-\mu}{\sigma}\right) = \\ &= \frac{1}{\sigma} f_Y\left(\frac{a-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(a-\mu)^2}{2\sigma^2}}. \end{aligned}$$

Definition A.3. ***(Log-normal distribution)***
The random variable $X > 0$ follows a Log-normal distribution $LN(\mu, \sigma^2)$ if its logarithm is Normal distributed $\log(X) \sim N(\mu, \sigma^2)$.

Example A.4. ***(Density of the Log-normal distribution)***
For $X \sim LN(\mu, \sigma^2)$ (and $Y = \log(X) \sim N(\mu, \sigma^2)$) we can write for the cumulative density function $F_X(a)$:

$$F_X(a) = \mathbb{P}[X \leq a] = \mathbb{P}[\log(X) \leq \log(a)] = F_Y(\log(a)).$$

This leads us to:

$$\begin{aligned} f_X(a) &= \frac{d}{da} F_X(a) = \frac{d}{da} F_Y(\log(a)) = \\ &= f_Y(\log(a)) \frac{1}{a} = \frac{1}{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\log(a)-\mu)^2}{2\sigma^2}} \end{aligned}$$

Definition A.5. ***(Chi-squared distribution)***
If $X_i \overset{\text{iid}}{\sim} N(0,1)$ for $1 \leq i \leq n$, then the random variable $Y = \sum_{i=1}^{n} X_i^2$ is Chi-squared distributed with n degrees of freedom. We will write $Y \sim \chi_n^2$ in this case.

Example A.6. ***(Density of the Chi-squared distribution)***
For $X \sim \chi_1^2$ (with $X = N^2, N \sim N(0,1)$) we can identify the density $f_X(a)$ as, by investigating the cumulative probability function $F_X(a)$:

$$\begin{aligned} F_X(a) &= \mathbb{P}[X \leq a] = \mathbb{P}[N^2 \leq a] \\ &= \mathbb{P}[-\sqrt{a} \leq N \leq \sqrt{a}] = F_N(\sqrt{a}) - F_N(-\sqrt{a}). \end{aligned}$$

This equation leads us to

$$\begin{aligned} f_X(a) = \frac{d}{da}F_X(a) &= \frac{d}{da}(F_N(\sqrt{a}) - F_N(-\sqrt{a})) \\ &= f_N(\sqrt{a})\frac{1}{2}a^{-1/2} - f_N(-\sqrt{a})\Big(-\frac{1}{2}\Big)a^{-1/2} \\ &= f_N(\sqrt{a})a^{-1/2} = \frac{1}{\sqrt{2\pi}}e^{-a/2}a^{-1/2}. \end{aligned}$$

Example A.7. ***(Distribution of $N(\mu, \sigma^2)$ sample variance 1)***
When we are looking at an random sample $X_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ for $1 \leq i \leq n$ the sample variance $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \hat{\mu}_n)^2$ is Chi-squared distributed with $n-1$ degrees of freedom, if $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$ denotes the sample mean. It is best shown with induction that

$$(n-1)\hat{\sigma}_n^2/\sigma^2 \sim \chi_{n-1}^2 \tag{A.2}$$

holds for $n \geq 2$:

- ***n=2:***

$$\begin{aligned} \hat{\sigma}_2^2/\sigma^2 &= \frac{1}{\sigma^2}\Big(X_1 - \frac{X_1+X_2}{2}\Big)^2 + \frac{1}{\sigma^2}\Big(X_2 - \frac{X_1+X_2}{2}\Big)^2 \\ &= \Big(\frac{X_1 - X_2}{2\sigma}\Big)^2 + \Big(\frac{X_2 - X_1}{2\sigma}\Big)^2 = \Big(\frac{X_1 - X_2}{\sqrt{2}\sigma}\Big)^2. \end{aligned}$$

  With small adoption to (A.7) *we see, that $X_1 - X_2$ is $N(0, 2\sigma^2)$ distributed and as a consequence $\frac{X_1 - X_2}{\sqrt{2}\sigma}$ is $N(0,1)$ distributed, hence its square follows a χ_1^2 distribution.*

- ***n >2:***
  We assume the equation holds for $n-1$ we will show it for n. Therefore we express the sample mean as:

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n}\Big(\sum_{i=1}^{n-1} X_i + X_n\Big) = \frac{1}{n}\Big((n-1)\hat{\mu}_{n-1} + X_n\Big),$$

and the sample variance as:

$$\begin{aligned}(n-1)\hat{\sigma}_n^2 &= \sum_{i=1}^{n}(X_i-\hat{\mu}_n)^2 = \sum_{i=1}^{n}(X_i^2+\hat{\mu}_n^2-2X_i\hat{\mu}_n)\\ &= \sum_{i=1}^{n}X_i^2+n\hat{\mu}_n^2-2\hat{\mu}_n\sum_{i=1}^{n}X_i\\ &= \sum_{i=1}^{n}X_i^2-n\hat{\mu}_n^2.\end{aligned}$$

This leads us to:

$$\begin{aligned}&(n-1)\hat{\sigma}_n^2\\ &= \sum_{i=1}^{n}X_i^2-n\hat{\mu}_n^2\\ &= \sum_{i=1}^{n-1}X_i^2+X_n^2-n\left(\frac{1}{n}((n-1)\hat{\mu}_{n-1}+X_n)\right)^2\\ &= \sum_{i=1}^{n-1}X_i^2-(n-1)\hat{\mu}_{n-1}^2+(n-1)\hat{\mu}_{n-1}^2+X_n^2\\ &-\frac{1}{n}\left((n-1)^2\hat{\mu}_{n-1}^2+X_n^2+2(n-1)\hat{\mu}_{n-1}X_n)\right)\\ &= (n-2)\hat{\sigma}_{n-1}^2+\frac{n-1}{n}\left(\hat{\mu}_{n-1}^2+X_n^2-2\hat{\mu}_{n-1}X_n\right)\\ &= (n-2)\hat{\sigma}_{n-1}^2+\frac{n-1}{n}\left(X_n-\hat{\mu}_{n-1}\right)^2,\end{aligned}$$

and by dividing through σ^2 we receive the equation:

$$(n-1)\hat{\sigma}_n^2/\sigma^2 = (n-2)\hat{\sigma}_{n-1}^2/\sigma^2+\left(\frac{X_n-\hat{\mu}_{n-1}}{\sigma\sqrt{n/(n-1)}}\right)^2.$$

Also here we can adopt (A.7) *to see, that $X_n-\hat{\mu}_{n-1}$ is $N(0,n/(n-1)\sigma^2)$ distributed and as a consequence $\frac{X_n-\hat{\mu}_{n-1}}{\sigma\sqrt{n/(n-1)}}$ is $N(0,1)$ distributed, hence its square follows a χ^2_{n-1} distribution. The needed independence is given according to (Rice, 2006, pp. 197 ff.).*

With the result $(n-1)\hat{\sigma}_n^2/\sigma^2 \sim \chi^2_{n-1}$ we know the distribution of $\hat{\sigma}_n^2$, unfortunately this representation depends on the unknown parameter σ.

Definition A.8. ***(Student's t-distribution)***
If $X \sim N(0,1)$ and $Y \sim \chi^2_{n-1}$ are independent random variables, then

$$Z = \frac{X}{\sqrt{Y/(n-1)}} \tag{A.3}$$

is Student's t-distributed with $n-1$ degrees of freedom. We will write $Z \sim T_{n-1}$ in this case.

Example A.9. ***(Distribution of*** $N(\mu, \sigma^2)$ ***sample variance 2)***
Student (1908) investigated the distribution of the random variable $\frac{\hat{\mu}_n - \mu}{\hat{\sigma}_n/\sqrt{n}}$*. Therefore it is rewritten as:*

$$\frac{\hat{\mu}_n - \mu}{\hat{\sigma}_n/\sqrt{n}} = \frac{\frac{\hat{\mu}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\hat{\sigma}_n^2/\sigma^2}}.$$

According to (A.8) *the expression* $\frac{\hat{\mu}_n - \mu}{\sigma/\sqrt{n}}$ *is* $N(0,1)$ *distributed. Moreover* $(n-1)\hat{\sigma}_n^2/\sigma^2 \sim \chi^2_{n-1}$ *holds due to* (A.2)*. This leads us in combination with* (A.3) *to a* T_{n-1} *distributed random variable, since* $\hat{\mu}_n$ *and* $\hat{\sigma}_n$ *are independent (Rice, 2006, pp. 195 ff.).*

Definition A.10. ***(Snedecor's F-distribution)***
For two independent random variables $X \sim \chi^2_n$ *and* $Y \sim \chi^2_m$ *the quotient:*

$$F = \frac{X/(n-1)}{Y/(m-1)}$$

follows a Snedecor's F-distribution $F_{n,m}$ *with* n *and* m *degrees of freedom.*

Example A.11. ***(Comparison of variances - normal distribution)***
For an iid sample $\boldsymbol{X}_n$ *of normal distributed variables* $X_i = N(\mu, \sigma^2), 1 \leq i \leq n$ *the following equation holds according to* (A.7)*:*

$$(n-1)\hat{\sigma}_X^2/\sigma^2 \sim \chi^2_{n-1}.$$

For two independent samples $\boldsymbol{X}_n$ *and* $\boldsymbol{X}_m$ *with* $X_i \overset{\text{iid}}{\sim} N(\mu_X, \sigma_X^2)$*,* $Y_i \overset{\text{iid}}{\sim} N(\mu_Y, \sigma_Y^2)$ *and estimates* $\hat{\sigma}_n$*,* $\dot{\sigma}_m$ *the quotient*

$$\frac{\hat{\sigma}_n^2/\sigma_X^2}{\dot{\sigma}_m^2/\sigma_Y^2}$$

is $F_{n-1,m-1}$ *distributed.*

Definition A.12. ***(Multivariate standard normal distribution)***
A random vector X *of dimension* n *is standard normal distributed, if its entries* $X_i \quad 1 \leq i \leq n$ *are standard normal distributed and independent, i.e.* $X_i \overset{\text{iid}}{\sim} N(0,1) \quad 1 \leq i \leq n$*. We write* $X \sim N(0, I_n)$*.*

Example A.13. ***(Density of the multivariate standard normal distribution)***
We start with determining the joint cumulative distribution function $F_X(a) : \mathbb{R}^n \to [0,1]$ *for a random vector* $X \sim N(0, I_n)$*:*

$$\begin{aligned} F_X(a) &= \mathbb{P}[X \leq a] = \mathbb{P}[X_1 \leq a_1 \wedge \cdots \wedge X_n \leq a_n] \\ &= \prod_{i=1}^n \mathbb{P}[X_i \leq a_i] = \prod_{i=1}^n F_{X_i}(a_i) = \prod_{i=1}^n \int_{-\infty}^{a_i} f_{X_i}(\tau_i) d\tau_i \\ &= \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} \prod_{i=1}^n f_{X_i}(\tau_i) d\tau_1 \cdots d\tau_n \end{aligned}$$

Now, we can calculate the density function $f_X(a)$ as:

$$\begin{aligned} f_X(a) &= \prod_{i=1}^{n} f_{X_i}(a_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-a_i^2/2} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n} a_i^2} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\frac{a^T a}{2}} \end{aligned}$$

Definition A.14. ***(Multivariate normal distribution)***
A random vector X of dimension n is normal distributed, if there exists a standard normal distributed random vector Z of dimension k, such that $X = AZ + \mu$ holds for $\mu \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times k}$. We write $X \sim N(\mu, AA^T)$.

Definition A.15. ***(Weibull distribution)***
The random variable X follows a Weibull distribution $Wbl(\lambda, k)$, if its density is given as:

$$f_X(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad \lambda > 0, k > 0,$$

for $x \geq 0$ and zero otherwise.

Definition A.16. ***(Gamma distribution)***
The random variable X follows a Gamma distribution $\Gamma(k, \theta)$, if its density is given as:

$$f_X(x) = \frac{x^{k-1} e^{-x/\theta}}{\Gamma(k)\theta^k}, \quad k > 0, \theta > 0,$$

for $x > 0$ and zero otherwise, with $\Gamma(\cdot)$ as the gamma function:

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

Definition A.17. ***(GEV distribution)***
The random variable X follows a Generalized Extreme Value (GEV) distribution distribution $GEV(\xi, \mu, \sigma)$, if its density is given as:

$$f_X(x) = \begin{cases} \frac{1}{\sigma}\left[1 + \gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{(-1/\xi)-1} e^{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}} & \xi \neq 0 \\ \frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} e^{-e^{-\frac{x-\mu}{\sigma}}} & \xi = 0 \end{cases}, \tag{A.4}$$

with $\xi \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0$ and $1 + \xi\frac{(x-\mu)}{\sigma} > 0$. The support is given with $x \in [\mu - \sigma/\xi, \infty)$ for $\xi > 0$, $x \in (-\infty, \mu - \sigma/\xi]$ for $\xi < 0$, and $x \in \mathbb{R}$ for $\xi = 0$.

A.2 Moment-generating function

The following section is based on the explanation of (Rice, 2006, pp. 155 ff.).

Definition A.18. ***(Moment-generating function)***
The moment generating function (MGF) for the continuous random variable X at time t is defined as

$$M_X(t) := \int_{-\infty}^{\infty} e^{tx} f_X(x)dx = \mathbb{E}[e^{tX}],$$

if this term exists.

Proposition A.19. ***(MGF under linear transformation)***
For a random variable X with MGF $M_X(t)$ we can express the MGF of $Y = a + bX$ with $a, b \in \mathbb{R}$ as

$$M_Y(t) = M_{a+bX}(t) = e^{at} M_X(bt). \tag{A.5}$$

Proof.

$$\begin{aligned} M_Y(t) &= \mathbb{E}(e^{tY}) = \mathbb{E}(e^{at+btX}) \\ &= \mathbb{E}(e^{at} e^{btX}) = e^{at}\mathbb{E}(e^{btX}) = e^{at} M_X(bt). \end{aligned}$$

□

Example A.20. ***(MGF for $Y \sim N(\mu, \sigma^2)$)***
For $X \sim N(0, 1)$ we obtain with (A.1) *for the moment generating function:*

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t^2/2-(x-t)^2/2} dx = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx \\ &= \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} du = e^{t^2/2}. \end{aligned}$$

This leads us for $Y = \mu + \sigma X \sim N(\mu, \sigma^2)$ in combination with (A.5) *to:*

$$M_Y(t) = e^{\mu t} M_X(\sigma t) = e^{\mu t} e^{\frac{\sigma^2}{2} t^2}.$$

Example A.21. ***(MGF for $X \sim \chi_1^2$)***
We examine the MGF for $X \sim \chi_1^2$, using the density function of the chi-square

distribution (see Example A.6):

$$\begin{aligned}
M_X(t) &= \int_0^\infty e^{tx} f_Y(x)dx = \int_0^\infty e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x/2} x^{-1/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{x(t-1/2)} x^{-1/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_0^{-\infty} e^{-u} \left(\frac{u}{1/2-t}\right)^{-1/2} \frac{du}{1/2-t} \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1/2-t}} \int_0^\infty e^{-u} u^{-1/2} du \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1/2-t}} \Gamma(1/2) \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1/2-t}} \sqrt{\pi} \\
&= \frac{1}{\sqrt{2}} \frac{1}{\sqrt{1/2-t}} = (1-2t)^{-1/2}
\end{aligned}$$

We used the substitution

$$\begin{aligned}
u &= x(1/2-t), \\
dx &= \frac{du}{1/2-t},
\end{aligned}$$

for which the integration boarders stay the same ($(1/2-t) > 0$ for interesting values $t < 1/2$).

Proposition A.22. ***(MGF of the sum of independent random variables)*** *For $X_i \quad 1 \leq i \leq n$ independent random variables the MGF of $Y = \sum_{i=1}^n X_i$ is given as:*

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t).$$

Proof.

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{t\sum_{i=1}^n X_i}] = \mathbb{E}[\prod_{i=1}^n e^{tX_i}] = \prod_{i=1}^n \mathbb{E}[e^{tX_i}] = \prod_{i=1}^n M_{X_i}(t)$$

□

Example A.23. ***(MGF for $X \sim \chi_n^2$)*** *Based on the MGF of a χ_1^2 random variable (see Example A.21), we can calculate the MGF of the random variable $X \sim \chi_n^2$. Let $X = \sum_{i=1}^n N_i^2$ for $N_i \overset{\text{iid}}{\sim} N(0,1)$, then the $N_i^2 \sim \chi_1^2$ are independent and we can write according to Proposition A.22:*

$$M_X(t) = \prod_{i=1}^n M_{N_i^2}(t) = M_{N_1^2}^n(t) = (1-2t)^{-n/2}$$

Definition A.24. ***(Moment-generating function of a random vector)*** *The moment generating function (MGF) for the random vector $X = (X_1, \cdots, X_n)$ at time $t = (t_1, \cdots, t_n)$ is defined as*

$$M_X(t) := \mathbb{E}[e^{t^T X}],$$

if this term exists.

Proposition A.25. ***(MGF under linear transformation - random vector)*** *For an n-dimensional random vector X with MGF $M_X(t)$ we can express the MGF of $Y = a + BX$ with $a \in \mathbb{R}^n$ and $B \in \mathbb{R}^{n \times n}$ as*

$$M_Y(t) = M_{a+BX}(t) = e^{a^T t} M_X(Bt). \tag{A.6}$$

Proof.

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{t^T Y}] = \mathbb{E}[e^{t^T(a+BX)}] = \mathbb{E}[e^{t^T a} e^{t^T BX}] \\ &= \mathbb{E}[e^{t^T a}]\mathbb{E}[e^{t^T BX}] = \mathbb{E}[e^{t^T a}]\mathbb{E}[e^{(B^T t)^T X}] \\ &= e^{t^T a} M_X(B^T t) \end{aligned}$$

□

Example A.26. ***(MGF for $Y \sim N(\mu, \Sigma)$)*** *For the random vector $X \sim N(0, I_n)$ we obtain with the density of the random vector (see Example A.13) the moment generating function:*

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{t^T X}] &= \prod_{i=1}^{n} \mathbb{E}[e^{t_i X_i}] = \prod_{i=1}^{n} M_{X_i}(t) \\ &= \prod_{i=1}^{n} e^{t_i^2/2} = e^{\frac{1}{2}\sum_{i=1}^{n} t_i^2} = e^{\frac{t^T t}{2}} \end{aligned}$$

Let now be $Y = AX + \mu$, with $\Sigma = AA^T$ and hence $Y \sim N(\mu, \Sigma)$, according to Definition A.14.

$$\begin{aligned} M_Y(t) = M_{AX+\mu}(t) &= e^{t^T \mu} M_X(A^T t) = e^{t^T \mu} e^{\frac{1}{2}(A^T t)^T A^T t} \\ &= e^{t^T \mu} e^{\frac{1}{2} t^T A A^T t} = e^{t^T \mu} e^{\frac{1}{2} t^T \Sigma t} \end{aligned}$$

The following theorem was for example stated by Rice (2006, pp. 155).

Theorem A.27.
If the moment generating function of X is finite for $t \to 0$ then the distribution of X is uniquely determined. In other words, this means for two random variables X and Y with $M_X(t) < \infty$ and $M_Y(t) < \infty$ for $t \in (-\epsilon, \epsilon)$ we can say:

$$P_X = P_Y \qquad \Leftrightarrow \qquad M_X(t) = M_Y(t) \quad \forall t \in (-\epsilon, \epsilon).$$

Example A.28. ***(Distribution of the sum - Normal distribution)***
When we are looking at a sequence of independent normal distributed random variables $X_i \sim N(\mu_i, \sigma_i)$ for $1 \leq i \leq n$ we can see, that the sum $Y = \sum_{i=1}^n X_i \sim N(\hat{\mu}, \hat{\sigma}^2)$ of this sequence is also normal distributed. Moreover we can show $\hat{\mu} = \sum_{i=1}^n \mu_i$ and $\hat{\sigma}^2 = \sum_{i=1}^n \sigma^2$:

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t) = \prod_{i=1}^n e^{\mu_i t} e^{\frac{\sigma_i^2}{2} t^2} = e^{\sum_{i=1}^n \mu_i t} e^{\frac{\sum_{i=1}^n \sigma_i^2}{2} t^2} = e^{\hat{\mu} t} e^{\frac{\hat{\sigma}^2}{2} t^2}. \tag{A.7}$$

When we are looking at $\bar{X} = \frac{Y}{n} = \frac{1}{n} \sum_{i=1}^n X_i$ and $X_i \overset{\text{iid}}{\sim} N(\mu, \sigma)$ we get in analogy to the above situation $\bar{X} \sim N(\bar{\mu}, \bar{\sigma})$ with $\bar{\mu} = \mu$ and $\bar{\sigma} = \frac{\sigma}{\sqrt{n}}$:

$$\begin{aligned} M_{\bar{X}}(t) &= \prod_{i=1}^n M_{X_i/n}(t) = \prod_{i=1}^n \mathbb{E}[e^{tX_i/n}] = \prod_{i=1}^n e^{\mu t/n} e^{\frac{\sigma^2}{2}(t/n)^2} \\ &= e^{\mu t} e^{\frac{(\sigma/\sqrt{n})^2}{2} t^2} = e^{\bar{\mu} t} e^{\frac{\bar{\sigma}^2}{2} t^2}. \end{aligned} \tag{A.8}$$

Example A.29. ***(Difference of independent χ^2 random variables)***
Let $X \sim \chi_p^2$ and $Y \sim \chi_q^2$ be two random variables with $p, q \in \mathbb{N}_{>0}$ and $p > q$. Then we can calculate the distribution of $Z = X - Y$ via the equation $X = Z + Y$, the Proposition A.22, and Example A.23, if Z and Y are independent as:

$$\begin{aligned} M_X(t) &= M_Z(t) M_Y(t) \\ (1-2t)^{-p/2} &= M_Z(t)(1-2t)^{-q/2} \\ M_Z(t) &= (1-2t)^{-(p-q)/2} \end{aligned}$$

Due to the unique characteristic of the MGF (see Theorem A.27) the random variable Z is also chi-square distributed with $(p-q)$ degrees of freedom, i.e. $Z \sim \chi_{p-q}^2$.

Example A.30. ***(Distribution of the sum - normal distribution vectors)***
When we are looking at independent random vectors $X_i \sim N(\mu_i, \Sigma_i) \quad i \leq k$ and their sum $Y = \sum_{i=1}^k X_i$, we obtain for its MGF:

$$\begin{aligned} M_Y(t) &= M_{\sum_{i=1}^k X_i}(t) = \mathbb{E}[e^{t^T \sum_{i=1}^k X_i}] \\ &= \prod_{i=1}^k \mathbb{E}[e^{t^T X_i}] = \prod_{i=1}^k e^{t^T \mu_i} e^{\frac{1}{2} t^T \Sigma_i t} \\ &= e^{t^T \sum_{i=1}^k \mu_i} e^{\frac{1}{2} t^T \sum_{i=1}^k \Sigma_i t} \end{aligned}$$

Hence, we obtain $Y \sim N(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \Sigma_i)$, i.e. the sum of independent normal distributed random vectors is itself a normal distributed random vector.

A.3 Linear Model - Proofs

Theorem A.31. ***(Least square estimate $\hat{\beta}$)***
The least square estimate in a Linear Model, minimizing the squared error term

$$\mathcal{S}(\beta) = \sum_{i=1}^{n} (y_i - \mu_i(\beta))^2 = \|y - X\beta\|^2,$$

is given with

$$\hat{\beta} = \arg\min_{\beta} \mathcal{S}(\beta) = (X^T X)^{-1} X^T y.$$

Proof.
We need to minimize the squared error sum $\mathcal{S}(\beta)$ with respect to β, which can be rewritten as:

$$\mathcal{S}(\beta) = \|y - X\beta\|^2 = (y - X\beta)^T \cdot (y - X\beta) = y^T y - 2\beta^T X^T y + \beta^T X^T X \beta.$$

By derivation of this term we obtain:

$$\frac{\partial}{\partial \beta} \mathcal{S}(\beta) = -2X^T y + 2X^T X \beta.$$

We achieve a potential minimum by setting this derivation to zero; For the the least square estimate $\hat{\beta}$ we obtain:

$$X^T X \hat{\beta} = X^T y.$$

For a regular matrix $X^T X$ we get the least square estimate $\hat{\beta}$ as

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

Since $\frac{\partial^2}{\partial \beta^2} \mathcal{S}(\beta) = 2X^T X$ we yield a positive semidefinite matrix, this term is indeed a minimum. □

Theorem A.32. ***(Distribution of $\hat{\beta}$)***
The least square estimate $\hat{\beta}$ in a Linear Model is normal distributed, with expectation β and variance $(X^T X)^{-1}\sigma^2$, i.e.:

$$\hat{\beta} \sim N(\beta, (X^T X)^{-1} \sigma^2).$$

Proof.
According to Example A.28 the estimate $\hat{\beta}$ as sum of normal distributed random vectors is also normal distributed. Furthermore, we see:

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}[(X^T X)^{-1} X^T y] = (X^T X)^{-1} X^T \mathbb{E}[y] = (X^T X)^{-1} X^T X \beta = \beta,$$

and

$$Var[\hat{\beta}] = Var[(X^T X)^{-1} X^T y] = (X^T X)^{-1} X^T Var[y] X (X^T X)^{-1} = (X^T X)^{-1} \sigma^2.$$

□

Theorem A.33. ***(Independence of $\hat{\beta}$ and $\mathcal{S}(\hat{\beta})$)***
In the Linear Model the estimates $\hat{\beta}$ and $\mathcal{S}(\hat{\beta})$ are independent.

Proof.
First we show, that $Cov(\hat{\beta}, r) = 0$, whereas $r = y - X\hat{\beta}$, and $\mathcal{S}(\hat{\beta}) = r^T r$. We rewrite $\hat{\beta}$ and r, using the hat matrix $H := X(X^TX)^{-1}X^T$ and its property:

$$(I - H)X = (I - X(X^TX)^{-1}X^T)X = X - X(X^TX)^{-1}X^TX = X - X = 0.$$

This leads us to:

$$\hat{\beta} = (X^TX)^{-1}X^Ty = (X^TX)^{-1}X^T(X\beta + \epsilon) = \beta + (X^TX)^{-1}X^T\epsilon,$$
$$r = y - X\hat{\beta} = y - X(X^TX)^{-1}X^Ty = (I - H)y = (I - H)(X\beta + \epsilon) = (I - H)\epsilon.$$

For the covariance matrix of $\hat{\beta}$ and r we get:

$$\begin{aligned} Cov(\hat{\beta}, r) &= Cov(\beta + (X^TX)^{-1}X^T\epsilon, (I - H)\epsilon) \\ &= (X^TX)^{-1}X^TCov(\epsilon, \epsilon)(I - H) \\ &= \sigma^2(X^TX)^{-1}X^T(I - H) \\ &= \sigma^2(X^TX)^{-1}X^T - \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = 0. \end{aligned}$$

Second, we show, that r is normal distributed. We can write $r = y - X\hat{\beta} = (I - H)y$ and according to Example A.28 it is normal distributed as linear combination of normal distributed random vectors.
Since both, $\hat{\beta}$ and r are normal distributed this implies, that they are independent. The same holds for $\hat{\beta}$ and $\mathcal{S}(\hat{\beta})$, since $\mathcal{S}(\hat{\beta}) = r^Tr$ can be written as function of r.
□

Theorem A.34. ***(Distribution of $\mathcal{S}(\hat{\beta})/\sigma^2$)***
In the Linear Model the random variable $\mathcal{S}(\hat{\beta})/\sigma^2$ is χ^2 distributed with $(n - p)$ degrees of freedom.

Proof.
When we are looking at $\epsilon^T\epsilon$, we can rewrite this term as:

$$\begin{aligned} \epsilon^T\epsilon &= (y - X\beta)^T(y - X\beta) \\ &= (y - X\hat{\beta} + X\hat{\beta} - X\beta)^T(y - X\hat{\beta} + X\hat{\beta} - X\beta) \\ &= (r + X\hat{\beta} - X\beta)^T(r + X\hat{\beta} - X\beta) \\ &= (r + X(\hat{\beta} - \beta))^T(r + X(\hat{\beta} - \beta)) \\ &= r^Tr + (\hat{\beta} - \beta)^TX^TX(\hat{\beta} - \beta). \end{aligned}$$

In the last step we used the following equation:

$$r^TX = ((I - H)y)^TX = y^T(I - H)X = 0.$$

This leads us to the equation

$$r^T r/\sigma^2 = \epsilon^T \epsilon/\sigma^2 - (\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta)/\sigma^2. \tag{A.9}$$

We know, that $\hat{\beta} \sim N(\beta, (X^T X)^{-1}\sigma^2)$ and hence $(\hat{\beta}-\beta)(X^T X)^{1/2}/\sigma = N(0, I)$ and according to Definition A.5 we get

$$(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta)/\sigma^2 \sim \chi^2_p,$$

since β got p entries. Furthermore, the term $\epsilon^2\epsilon/\sigma^2$ can be rewritten as:

$$\epsilon^T \epsilon/\sigma^2 = \sum_{i=1}^{n} (\epsilon_i/\sigma)^2 \sim \chi^2_n.$$

Since $r^T r/\sigma^2$ and $(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta)/\sigma^2$ are independent (r and $\hat{\beta}$ are independent, see Theorem A.33) we know, that the distribution of $r^T r/\sigma^2$ is given as chi square distribution with $(n-p)$ degrees of freedom (see Example A.29), i.e. $r^T r/\sigma^2 = \mathcal{S}(\hat{\beta})/\sigma^2 \sim \chi^2_{n-p}$. □

Proposition A.35. ***(Hypothesis test for** $\beta_i = 0$**)***
The random variable $T = \frac{\hat{\beta}_i}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_i}}}$ *can be used to test the hypothesis* $H_0 : \beta_i = 0$. *The random variable* $\hat{\sigma}^2_{\hat{\beta}_i}$ *describes the* i*-th diagonal element of the matrix* $\hat{\sigma}^2(X^T X)^{-1}$. *Under* H_0 $T \sim t_{n-p}$ *holds.*

Proof.
Examining the term $\frac{\hat{\beta}_i-\beta_i}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_i}}}$ we can rewrite it with $\hat{\sigma}^2_{\hat{\beta}_i} = \hat{\sigma}^2 v_{i,i}$ ($\sigma^2_{\hat{\beta}_i} = \sigma^2 v_{i,i}$), whereas $v_{i,i}$ describes the i-th diagonal element of $(X^T X)^{-1}$ as:

$$\frac{\hat{\beta}_i-\beta_i}{\sqrt{\hat{\sigma}^2_{\hat{\beta}_i}}} = \frac{\frac{\hat{\beta}_i-\beta_i}{\sqrt{\sigma^2_{\hat{\beta}_i}}}}{\sqrt{\frac{n-p}{\sigma^2}\hat{\sigma}^2/(n-p)}} = \frac{\frac{\hat{\beta}_i-\beta_i}{\sqrt{\sigma^2_{\hat{\beta}_i}}}}{\sqrt{\frac{\mathcal{S}(\hat{\beta})}{\sigma^2}/(n-p)}} \sim T_{n-p}.$$

This is true according to Definition A.8, since $\frac{\hat{\beta}_i-\beta_i}{\sqrt{\sigma^2_{\hat{\beta}_i}}} \sim N(0,1)$ and $\frac{\mathcal{S}(\hat{\beta})}{\sigma^2} \sim \chi^2_{n-p}$ (see Theorem A.34) are independent due to the independence of $\hat{\beta}$ and $\mathcal{S}(\hat{\beta})$ (see Theorem A.33). □

Theorem A.36. ***(F-ratio hypothesis test)***
Having a model with p *parameters and an estimate* $\hat{\beta}$ *one can test* $q \leq p$ *parameters, when calculating the estimate* $\dot{\beta}$ *for the original model excluding the specified parameters. The test statistic is given with:*

$$F(\hat{\beta}, \dot{\beta}) = \frac{n-p}{q} \cdot \frac{\mathcal{S}(\dot{\beta}) - \mathcal{S}(\hat{\beta})}{\mathcal{S}(\hat{\beta})}.$$

Under H_0 $F(\hat{\beta}, \dot{\beta}) \sim F_{q,n-p}$ *holds.*

Proof.
See Wood (2006). □

B. References

Albert, W. A. J. (1838). Über Treibseile am Harz. *Archive für Mineralogie, Geognosie, Bergbau und Hüttenkunde*, 10:215–234.

Basquin, O. (1910). The exponential law of endurance tests. In *Proc. ASTM*, volume 10, Part II, page 625.

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18.

Draper, N. R. and Smith, H. (2014). *Applied regression analysis*. John Wiley & Sons, New York.

Ewing, J. A. and Rosenhain, W. (1899). Experiments in micro-metallurgy: effects of strain. preliminary notice. *Proceedings of the Royal Society of London*, 65(413-422):85–90.

Franke, J., Härdle, W. K., and Hafner, C. M. (2015). *Statistics of financial markets: an introduction*. Springer, New York.

Freudenthal, A. (1952). Planning and interpretation of fatigue tests. In *Symposium on Statistical Aspects of Fatigue*. ASTM International, Atlantic City.

Irwin, G. R. (1997). Analysis of stresses and strains near the end of a crack traversing a plate. *Spie Milestone series MS*, 137(167-170):16.

Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012). *Introduction to linear regression analysis*, volume 821. John Wiley & Sons, New York.

Nelder, J. A. and Baker, R. J. (1972). Generalized linear models. *Encyclopedia of statistical sciences*, 135(3):370–384.

Nordstokke, D. W. and Zumbo, B. D. (2010). A new nonparametric Levene test for equal variances. *Psicologica: International Journal of Methodology and Experimental Psychology*, 31(2):401–430.

Paris, P. C., Gomez, M. P., and Anderson, W. E. (1961). A rational analytic theory of fatigue. *The Trend in Engineering*, 13(1):9–14.

Pearson, S. (1975). Initiation of fatigue cracks in commercial aluminium alloys and the subsequent propagation of very short cracks. *Engineering Fracture Mechanics*, 7(2):235–247.

Rice, J. (2006). *Mathematical statistics and data analysis*. Cengage Learning, Duxbury.

Rüschendorf, L. (2014). *Mathematische Statistik*. Springer, Berlin Heidelberg.

Schlottmann, D. and Schnegas, H. (2016). *Auslegung von Konstruktionselementen: Sicherheit, Lebensdauer und Zuverlässigkeit im Maschinenbau*. Springer, Berlin Heidelberg.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611.

Shen, C.-l. (1994). The statistical analysis of fatigue data. *PhD thesis, The University of Arizona*.

Student (1908). The probable error of a mean. *Biometrika*, 6(1):1–25.

Suresh, S. (1998). *Fatigue of materials*. Cambridge University Press, Cambridge.

Suresh, S. and Moss, C. (1980). Near-threshold fatigue crack growth in 21/icr-1mo pressure vessel steel in air and hydrogen. *Journal of Engineering Materials and Technology*, 102:293–299.

Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC Press, Boca Raton.