TU Kaiserslautern **Fachbereich Mathematik**

Dissertation

Partially Passed Component Counting for Evaluating Reliability

Dipl.-Math. Sascha Feth

Gutachter:
Prof. Dr. Jürgen Franke, Technische Universität Kaiserslautern
Prof. Dr. Jacques de Maré, Chalmers University of Technology

Vom Fachbereich Mathematik der Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.)
genehmigte Dissertation

Datum der Disputation: 18. Dezember 2009

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

D386

Acknowledgement

My thanks go Prof. Dr. Jürgen Franke for the possibility to write my dissertation at his research group as well as his support and encouragement. At ITWM I want to thank first of all Dr. Klaus Dreßler and Dr. Michael Speckert for introducing me to the practical aspects of the topic and offering me a Phd position at ITWM. Thanks to Prof. Dr. Jacques de Maré for his willingness of beeing my co-referee. Finally thanks to all colleagues at ITWM and university for creating a nice working environment.

Dedicated Inge Weißmann

Contents

Chapter 1

Introduction, Overview and Notation

1.1 Objective: Homologation of Safety Relevant Components

Deciding about homologation of safety relevant components from passenger or utility cars means to verify reasonable large lifetimes under specified loads. Scattering production processes require to model the lifetime as a random variable $T \sim F$. Therefore the homologation claim can be formulated as:

> Survival probability at design/target life q_0 (**reliable life**) is greater than R_0 (**reliability**)

If q_{1-R_0} denotes the $(1 - R_0)$-quantile of the lifetime distribution F, then an equivalent formulation of the claim is:

$$q_0 < q_{1-R_0} \tag{1.1}$$

There is a popular approach referred to as **success run**:

> Grant homologation if N tested units reach lifetimes $L > q_0$, without exception.

The **experimental design** (N, L) is calculated such that the resulting statistical test will reach a desired significance α. In a generalised version of the success run ("success run with failures") it is allowed to observe some failures before time L, if only (N, L) are chosen large enough. There is even a Bayesian extension of the method, allowing to model prior knowledge with beta distributions as a conjugated family.

Unfortunately, the method shows two major drawbacks:

1. For fixed significance, the power of the test does depend strongly on the experimental design (N, L).

2. Success runs with failures do not allow for beta distributions as conjugated families.

The scope of this thesis is to further generalise the success run method to fix both drawbacks as far as possible by introducing **partially-passed component counting**.

In the remainder of this chapter a short introduction to the topic is given, before we start to analyse the approaches in detail.

1.2 Current Approach

1.2.1 Data Situation

Samples are drawn by testing several prototypes on test tracks or test rigs. Limiting the available sample sizes and/or the testing times leads to different **censoring patterns**:

- **Failure-censored sampling:** After r observed failures out of N tested units, the remaining $N-r$ experiments are suspended.
- **Time-censored sampling:** Suspend every item exceeding testing time L. The number of observed failures is random, but the maximal overall test time is known to be $N \cdot L$.

Depending on the censoring pattern a statistical method is chosen:

- **Reliability estimation test plans (RET)** try to observe as many failures as possible, for fitting parametric models (failure-censored sampling).
- **Reliability demonstration test plans (RDT)** try to demonstrate a minimum lifetime, working for time-censored sampling.

In fatigue studies time-censored sampling is the preferred censoring pattern and will be the focus of this thesis. Special attention is paid to the emerging small sample sizes, using the following paradigm:

P1 Information on the reliability at time q_0 shall be gathered by increasing the test duration L rather than the sample size N.

Every manufacturer possesses experience in developing safety relevant components. It is desired to use this experience in a quantitative way:

P2 The applied statistical test should allow for Bayesian methods.

1.2.2 Reliability Estimation using Quantile Estimation

Let $\hat{q}_{1-R_0}$ be a point estimate of the quantile q_{1-R_0}. To judge if (1.1) is satisfied, we compare the lower bound $\hat{q}_{1-R_0,\alpha}$ of a one-sided $(1-\alpha)$-cofidence interval (CI) to q_0. Two methods were chosen for further study:

- **Delta method:** Assume that $\hat{q}_{1-R_0}$ follows a normal distribution. The **Fisher information matrix** delivers the estimator's asymptotic covariance: $\mathbb{C}\text{ov}\left(\hat{\theta}\right) = (NI_N)^{-1}$. **Gaussian error propagation** provides estimates V_{1-R_0} for $\mathbb{V}\text{ar}(\hat{q}_{1-R_0})$. Using an asymptotic normal distribution for $\hat{q}_{1-R_0}$ gives confidence intervals for q_{1-R_0}:

$$\hat{q}_{1-R_0,\alpha} = \hat{q}_{1-R_0} - \sqrt{V_{1-R_0}} \cdot t_{N-1}^{-1}(1-\alpha)$$

This approach is motivated by asymptotic arguments and is algorithmically cheap.

- **Bootstrap methods:** Estimate the sampling distributions parameter by $\hat{\theta}$ to generate M samples of length N from $F_{\hat{\theta}}$. For each resample $\underline{x}^*$ the ML quantile estimate $\hat{q}^*_{p_0}$ is calculated. The (empirical) CDF G^* of $\hat{q}^*_{p_0}$ is an estimate for the CDF G of the estimator $\hat{q}_{p_0}$. There are two main types of bootstrap methods: **bootstrap quantile method (BQM)** and **bootstrap-t**, explained in section 2.3.

One exemplary simulation study about CI for means can be found in [19], showing good results for the bootstrap-t. Adapting it to censored survival data gives the well known **hybrid bootstrap**[1].

1.2.3 Reliability Demonstration using Success Runs

Success runs claim, that N units have to survive at least L cycles to achieve homologation (time-censored sampling). To verify a reasonably small failure quota $p = F(q_0) < p_0 = 1 - R_0$, count the number S_L of tested components $T_1, \ldots, T_N$ surviving time $L = \lambda q_0$, $\lambda \geq 1$. Reject $H_0 : p \geq p_0$ if $S_L = N$ (no failure before time L). If p_L denotes the failure probability at time L, then the binomial distribution of S_L gives the **significance equation**:

$$(1 - p_L)^N \leq \alpha \tag{1.2}$$

where α is the significance of the statistical test. It has to be assumed, that $p_0 = F(q_0)$ uniquely determines F within the chosen family, allowing the calculation of $p_L = F(L)$.

1.2.4 Bayesian Reliability Demonstration

Developing safety relevant components is rather evolutionary than revolutionary. To respect previous knowledge Bayesian statistics is used. If the **reliability** $R = 1 - p \in [0, 1]$ is used to describe quality, priors can be formulated and updated using the sample information. The theory of Bayesian reliability analysis is well developed, including methods for success runs. It can be shown, that binomial sampling with prior knowledge modeled by beta PDF $\text{Beta}(A_0; B_0)$ does lead to beta PDF for posterior knowledge:

For $L = q_0$ binomial sampling gives likelihoods proportional to

$$R^{S_L}(1 - R)^{N - S_L},$$

where $F(q_0) = 1 - R$. If a **beta distribution**

$$\pi \propto R^{A_0 - 1}(1 - R)^{B_0 - 1}$$

is used as a prior, then the posterior will be proportional to

$$R^{A_0 + S_L - 1}(1 - R)^{B_0 + N - S_L - 1}$$

[1] as explained in [16]

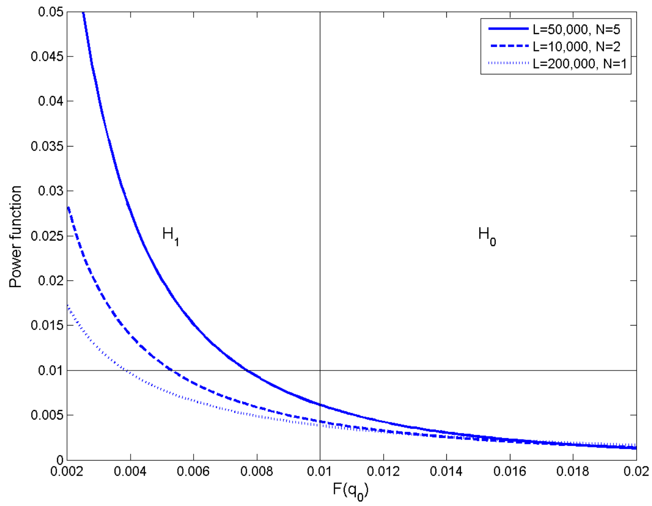

Figure 1.1: Test power for different (N, L), all fulfilling equation (1.2).

The result is a parameter update formula

$$A = A_0 + S_L,\; B = B_0 + N - S_L$$

It follows, that **beta distributions are conjugated to binomial sampling for** $L = q_0$.

1.2.5 Drawbacks and Gaps of the Current Approaches

Reliability Estimation

Small sample sizes will lead to large **coverage errors** of all CI. When used for testing they will give wrong significances. Time-censoring with small L might lead to completely censored samples. In this situation all quantile estimates will degenerate and one has to switch to reliability demonstration.

Reliability Demonstration

Success runs are well adapted to completely censored samples, but have major drawbacks concerning their power function. For given α there are countably many pairs (N, L) fulfilling equation (1.2). These pairs do **not** lead to the same **power of the test**, as figure 1.1 indicates. Small N are desirable to ensure affordability but decrease the probability of homologation for actually reliable designs.

Bayesian Reliability Demonstration

Since for Weibull distributions $(1 - F(L)) = (1 - F(q_0))^{\lambda^\gamma}$, where $\lambda = L/q_0$, the case $L > q_0$ does lead to a likelihood $R^{\lambda^\gamma \cdot S_L}(1 - R^{\lambda^\gamma})^{N - S_L}$. If $S_L < N$, then it is not possible to name the beta parameters of the posterior. This means, that one **runs the risk to have no parameterisation of the posterior** for $L > q_0$ and $S_L < N$.

1.3 Generalising Success Runs

The main topic of this thesis is to introduce a method working for all types of incomplete samples and power invariant w.r.t. N, L for given α.

1.3.1 Combining RET and RDT

Success runs are counting passed components in a discrete way:

> Every component T_i surviving time L will add summand 1 to S

Homologation is granted only if $S \geq S_{\text{crit}} \in \mathbb{N}$. Chapter 5 will modify the counting:

> Components T_i failing before time L will add a summand $s = s(T_i)$, where s is monotonically increasing with $s(0) = 0$ and $s(L) = 1$.

If we choose

$$s(T_i) = \frac{1}{p_L} F_{H_0}(T_i),\ B = \sum_{i=1}^{N} s(T_i)$$

then B (called **PPC-count**) is a direct generalisation of the success runs test statistic S. If the critical value for B is calculated correctly, the continuous nature of B allows the test to have **correct significance for every feasible pair** (N, L).

While the count is derived by generalising a reliability demonstration method, it also allows point estimation of the reliability for Weibull models. The resulting estimator in section 6.2 is equivalent to MLE for the sampling distribution. **In this sense PPC counting is some hybrid of RET and RDT**.

1.3.2 Applying Bayesian Methods

If the success run S_L is replaced by the PPC count B_L a beta PDF might still be used to get an easy knowledge-update formula. An approximation formula for the true posterior will be derived in section 7.2.2.

1.4 Overview

Censoring patterns determine which statistical method to use: Quantile estimates require a sufficient large number of failures, otherwise success runs are needed. **The main result will be, that for small time-censored samples the PPC count should be the favored method**. This is shown in several steps:

Chapter 2: Different algorithms for quantile-CI. In terms of their coverage error (deviation between nominal and empirical confidence) the asymptotic delta method is compared to some resampling methods. Both methods show inadmissible large coverage errors.

Section 2.5.1: Monte Carlo corrections of the delta method for lowering coverage errors. Even after correction the method will show undesirable discontinuities.

Chapter 3: Verification of $F(q_0) < p_0$ using binomial sampling to derive success runs. For small sample size, reliable designs are frequently refused (error of second kind).

Chapter 4: Randomisation of success runs by randomly branching the tests decision rule. The probability for errors of second kind decreases, but the problem of repeatability occurs.

Chapter 5: Introduction of partially passed component counting (PPC counting) as a generalisation of success runs. Comparing both concepts to randomisation makes the PPC counting the method to be preferred.

Chapter 6: Maximum likelihood estimation for CUS models are equivalent to MLE of the sampling distribution.

Appendix D Introduction to Bayesian statistics and conjugated families.

Chapter 7: Bayesian quality control for success runs and PPC-counts using beta distributions. The beta PDF will not be exactly conjugated to CUS sampling, but the approximation is reasonable. Different to binomial sampling the parameter update formulas are also available for $L > q_0$.

Chapter 8: Statistical tests coming from PPC counting are almost equivalent to corrected quantile estimates in terms of power functions, but are free of discontinuities.

1.5 Notation

Let $T \sim F$ with $F \in \{F_\theta \mid \theta \in \mathbb{R}^k\}$ (typically $k \in \{2,3\}$). Possible families for F are Weibull or lognormal (see appendix A). In this thesis, most results are formulated for Weibull distributions.

The homologation claim is formulated as a quantile hypothesis ,

$$\mathbb{P}(T < q_0) = F(q_0) < p_0 = 1 - R_0, \tag{1.3}$$

making $H_0 : F(q_0) \geq p_0$ the null hypothesis to be rejected. A sample with lifetimes $T_1, \ldots, T_N$ is drawn, using time-censoring.

Definition 1:
The censored version X^+ of a random variable X w.r.t. censoring time L is defined as:

$$X^+ = \begin{cases} X, & X \leq L \\ L, & X > L \end{cases}$$

*Further define the **censoring indicator** δ by:*

$$\delta = \begin{cases} 1, & X \leq L \\ 0, & X > L \end{cases}$$

*Samples with $\sum \delta = N$ are called **complete sample**.*

Statistical inference may only use the information available from $\underline{T}^+ = (T_1^+, \ldots, T_N^+)$ and $\underline{\delta} = (\delta_1, \ldots, \delta_N)$.

Prior knowledge about the reliability R will be described by a **prior density** $\pi(R)$. Togheter with the models likelihood $L(R \mid \underline{T}, \underline{\delta})$ a **posterior density** $\pi(R \mid \underline{T}, \underline{\delta})$ will be calculated using bayesian theorem:

$$\pi(R \mid \underline{T}, \underline{\delta}) \propto \bar{\pi} \cdot L(R \mid \underline{T}, \underline{\delta})$$

As a short form write $\bar{\pi}(R)$ for $\pi(R \mid \underline{T}, \underline{\delta})$, which should not be confused with the survivor function of π, which will be written as $1 - \Pi$ if it is needed.nMain candidate for π will be the **beta distribution** as introduced in section A.4.

Chapter 2

Quantile Estimation for Reliability Estimation

Consider the p_0-quantile q_{p_0} of F_θ from the homologation claim of section 1.1. Based on a sample $\underline{T}^+$, $\underline{\delta}$ a $(1-\alpha)$-CI $\hat{q}_{p_0,\alpha}$ is calculated using delta and bootstrap methods. In contrary to success runs (chapter 3) the exact lifetimes are needed, giving this approach the name **variables life test**. We start with looking at point estimates.

2.1 Parametric Point Estimates for Quantiles

Fatigue applications most frequently make use of **least squares in probability paper**. Details of the approach can be found in [2]. It is known, that this method has less efficiency than **maximum likelihood estimation** (MLE) (see [21]), which will be the preferred method here.

For incomplete samples the likelihood is defined to be:

$$L(\theta \mid \underline{x}, \underline{\delta}) = \prod_{\delta_i=1} f_\theta(x_i) \cdot \prod_{\delta_i=0} (1 - F_\theta(x_i)), \tag{2.1}$$

using the notation of definition 1. This modification ensures the consistency of the MLE in the presence of censored data, if the censoring pattern fulfills some independence conditions (see [20]). All patterns used in this thesis fulfill these conditions.

The argument $\hat{\theta}$, maximizing the log-likelihood $l = \ln L$, is used to construct the ML quantile estimate:

$$\hat{q}_p = F_{\hat{\theta}}^{-1}(p)$$

Remark 1 (Completely censored samples):
For completely censored samples the likelihood degenerates to:

$$L(\theta \mid \underline{x}) = \prod_{\delta_i=0} (1 - F_\theta(x_i))$$

E.g. think of a normal distribution with unknown μ. Obviously, L is monotonically increasing in μ, making it impossible to define the MLE for parameters or quantiles. This problem is addressed again in section 2.5.2.

2.2 Delta Method

The **delta method** derives quantile CI by writing point estimates as functions of the parameter estimates. This is done in 4 steps:

1. Establish an asymptotic normal distribution for the parameter estimate $\hat{\theta}$.

2. Get the asympotic covariance of $\hat{\theta}$ from the **Fisher information matrix**.

3. Use Gaussian error propagation for the **quantile function** to get the asymptotic variance of $\hat{q}_p$.

4. Finish with an estimated normal distribution for the asymptotically unbiased quantile estimator $\hat{q}_p$.

Details of the approach and corresponding proofs can be found in [10] or [11].

2.2.1 Theoretical Background

In this section, we review the standard theory of maximum likelihood estimates (MLE) without going into technical details. The regularity conditions which we skip in formulating the main results are typically used in particular in the situation of interest for this thesis satisfied. For the details we refer to the literature, e.g. [13].

Asymptotic Covariance and Fisher Information

Under some smoothness conditions most MLE $\hat{\theta}$ follow asymptotically a normal distribution:

$$\sqrt{N}\left(\hat{\theta}-\theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0;\Sigma) \tag{2.2}$$

To calculate the asymptotic covariance Σ, the concept of Fisher information is needed.

Definition 2:
*The **Fisher information matrix** is defined as the score function's second moment:*

$$\mathcal{I}_N(\theta) = \mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta_i} l(\underline{X} \mid \theta)\right)\left(\frac{\partial}{\partial\theta_j} l(\underline{X} \mid \theta)\right)\right)_{i,j}$$

If the model is correctly specified and the distribution is smooth enough as a function of θ, then $\mathcal{I}(\theta)$ may also be written as:

$$\mathcal{I}_N(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta_i\theta_j} l(\underline{X} \mid \theta)\right)$$

Theorem 1
Under some smoothness conditions, it holds that:

$$\sqrt{N}\left(\hat{\theta}-\theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}_p\big(0;\mathcal{I}_1(\theta)^{-1}\big) \tag{2.3}$$

Delta Method and Error Propagation

Theorem 2 (Delta Method)
Let $X_n \in \mathbb{R}^p$ be an asymptotic normal random vector with unit covariance matrix

$$a_n (X_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0; I_p)$$

Further $g : \mathbb{R}^p \to \mathbb{R}^q$ be a mapping, differentiable at μ, then:

$$a_n(g(X_n) - g(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}_q\big(0; \nabla g(\mu) \cdot \nabla g(\mu)^t\big)\,,$$

Proof *Perform a Taylor expansion of g around μ:*

$$g(X_n) = g(\mu) + \nabla g(\mu)(X_n - \mu) + \textit{higher order terms}$$

Since X_n converges in probability to μ, the remaining higher order terms converge in probability to zero. Finally, applying Slutsky's theorem to

$$a_n(g(X_n) - g(\mu)) = \nabla g(\mu)\underbrace{a_n(X_n - \mu)}_{\to \mathcal{N}_p(0;I_p)} + \underbrace{a_n(\textit{higher order terms})}_{\to 0}$$

gives the desired result.

qed

Using Theorem 2 for a function g of the parameter estimates gives:

$$\begin{aligned} \sqrt{N}\left(\hat{\theta} - \theta\right) &\xrightarrow{\mathcal{L}} \mathcal{N}_p(0; \Sigma) \\ \Rightarrow \sqrt{N}\left(g(\hat{\theta}) - g(\theta)\right) &\xrightarrow{\mathcal{L}} \mathcal{N}_p\big(0; \nabla g(\mu)\Sigma\nabla g(\mu)^t\big) \end{aligned} \tag{2.4}$$

2.2.2 Estimation of the Asymptotic Covariance

For i.i.d. data $\underline{X} = (X_1, \ldots, X_N)$ we have immediately

$$l(\theta \mid \underline{X} = \sum_{i=1}^{N} l(\theta \mid X_i)$$

and therefore $\mathcal{I}_N(\theta) = N\mathcal{I}_1(\theta)$. Definition 2 has given two interpretations of $\mathcal{I}_N(\theta)$: Second likelihood derivatives (under regularity conditions) or squared score function (general case). Therefore two estimators of $\mathcal{I}_1(\theta)$ may be defined:

Definition 3 (Hesse estimator):

$$\hat{\mathcal{I}}_{\text{Hesse}}(\theta) = \left(-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2}{\partial\theta_i\partial\theta_j} l(\theta \mid X_i)\right)_{i,j\in\{1,\ldots,k\}} \tag{2.5}$$

Definition 4 (BHHH estimator (Bernd-Hall-Hall-Hausmann)):

$$\hat{\mathcal{I}}_{\text{BHHH}}(\theta) = \frac{1}{N}\left(\sum_{i=1}^{N} \nabla l(\theta|X_i) \cdot \nabla l(\theta|X_i)^t\right) \tag{2.6}$$

Finally estimate Σ by:

$$\hat{\Sigma} = \hat{\mathcal{I}}_1(\theta)^{-1}$$

2.2.3 Application to Quantile CI

To use equation (2.4) for quantile CI estimation consider the quantile function $g(\cdot) = F_{\cdot}^{-1}(p)$, $\theta \in \Theta \subset \mathbb{R}^2$ and get estimates for $g(\theta) = q_p$ and Σ.

The asymptotic normal distribution of $\hat{q}_p$ is now used to construct a one-sided CI:

$$\boxed{q_p \leq \hat{q}_{p,\alpha} = \hat{q}_p + t_{N-1}^{-1}(1-\alpha) \cdot \hat{\sigma}_{\hat{\theta}}, \text{ with probability } \alpha} \tag{2.7}$$

(Usage of a student distribution to take into account the estimated variance)

Remark 2 (Logarithmic Delta Method):
Appendix B gives the necessary formulas for censored Weibull samples. Logarithmic transformations will be used to work with Gumbel distributions, being closer to a normal distribution than the original Weibull distribution. Further details will be explained in section 2.3.5, when using logarithmic transformations for bootstrap methods.

2.2.4 Coverage error

The derivation of the delta CI (2.7) has used several assumptions:

1. The inverse Fisher information approximates the covariance matrix Σ.
2. Delta method can be used to get the variance of $\hat{q}_p$.
3. $\hat{q}_p$ is asymptotically normally distributed.
4. $\hat{\Sigma}$ was a good estimate for Σ.
5. $\hat{q}_p$ is asymptotically unbiased.

Definition 5:
Let I be any $(1-\alpha)$-confidence interval for the unknown true value $g(\theta)$, then $\mathbb{P}(g(\theta) \in I)$ is called coverage probability or empirical confidence of I. $1-\alpha$ is called nominal confidence. The deviation $|\mathbb{P}(g(\theta) \in I) - (1-\alpha)|$ will be called coverage error.

For small samples the upper assumptions will not hold, making the delta methods coverage error large (see section 2.4). Resampling methods are known[1] to work quite well for small samples and will be studied in the next section.

Additionally, assumption 1 is crucial. As, due to the small sample sizes in the applications we are interested in, we have to restrict ourselves to simple parametric models, we expect those models to be misspecified. In that case, the two expressions which we have in Definition 2 do not coincide, an the asymptotic covariance matrix of $\hat{\theta}$ is of a more complicated form. This could also be estimated consistently refering to the law of large numbers, but as asymptotic normality does not provide a good enough approximation for small sample anyhow, we do not go into that direction.

[1] see e.g. [19]

2.3 Bootstrap Methods

Two main types of bootstrap methods are presented: Bootstrap-t and bootstrap quantile method. Adapting both concepts to censored data will lead to the hybrid bootstrap. A complete discussion can be found in [16] or [15].

2.3.1 Bootstrap Quantile Method

Consider a statistic $T = T(\underline{X}) \sim G$, estimating $g(\theta)$. Let $\hat{F}$ estimate the sampling distribution F (empirical-CDF or MLE). For each resample $\underline{X}^*$, drawn from $\hat{F}$, the statistic $T^* = T(\underline{X}^*) \sim G^*$ can be calculated. Repeating this multiple times, say M, the empirical CDF of $T_1^*, \dots, T_M^*$ approximates G^* (we will not distinguish between G^* and its step function approximation).

The **bootstrap quantile method (BQM)** intuitively uses G^*-quantiles to construct an equal-tailed CI for $g(\theta)$:

$$\boxed{\mathbb{P}\left(G^{*,-1}\left(\frac{\alpha}{2}\right) \leq g(\theta) \leq G^{*,-1}\left(1-\frac{\alpha}{2}\right)\right) \approx 1-\alpha} \tag{2.8}$$

The true tail probabilities are not equal, since in general $G^* \neq G$. To correct for the resulting coverage errors, there is a bias corrected version of the BQM.

2.3.2 Bias Corrected Bootstrap Quantile Method

To motivate the bias corrected BQM the concept of pivotal statistics is needed:

Definition 6:
A random variable $T = T(\underline{X}) \sim G$, $X \sim F \in \{F_\theta \mid \theta \in \Theta \subset \mathbb{R}^k\}$ is called pivotal, if G does not depend on θ. E.g. the distribution of the familiar t-statistic does not depend on the parameter of the gaussian data.

Assume, that there is a $z_0, \sigma \in \mathbb{R}$ and a monotonically transformation s, making $S(\underline{X})$ pivotal[2]:

$$\begin{aligned} s(T) - s(g(\theta)) &\sim \mathcal{N}\left(-z_0\sigma; \sigma^2\right), \\ s(T^*) - s(T) &\sim_* \mathcal{N}\left(-z_0\sigma; \sigma^2\right) \end{aligned}$$

Consequently, a CI for $s(g(\theta))$ can be defined by:

$$s(T) + z_0\sigma \pm \sigma\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$$

z_0 is yet unknown, but using s^{-1} on the equation

$$\mathbb{P}_*(s(T^*) \leq s(T)) = \Phi(z_0)$$

gives:

$$z_0 = \Phi^{-1}\left(G^*(T^*)\right), \; T^* \sim G^*$$

[2]Compare: hybrid bootstrap in section 2.3.4

Definition 7 (Bias corrected BQM):

$$\left[G^{*,-1}\left(\Phi\left(2z_0 - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)\right),\, G^{*,-1}\left(\Phi\left(2z_0 + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)\right)\right] \quad (2.9)$$

As a special case $z_0 = 0$ is giving the ordinary BQM:

$$\left[G^{*,-1}\left(\frac{\alpha}{2}\right),\, G^{*,-1}\left(1-\frac{\alpha}{2}\right)\right] \quad (2.10)$$

Changing the lower normal quantile gives left-sided CI:

$$\boxed{\left[G^{*,-1}\left(\Phi\left(2z_0 - \Phi^{-1}(1-\alpha)\right)\right),\, \infty\right)} \quad (2.11)$$

2.3.3 Bootstrap-t Method

Since F is substituted by $\hat{F}$, the dependence of G on F should be minimal.

Bootstrap-t applies BQM to pivotal statistics. Since pivotal transformation are seldom available, studentisation is used:

$$S = \frac{T - g(\theta)}{\hat{\sigma}}$$

$\hat{\sigma}$ is the estimated[3] variance of T. Due to replacing σ by $\hat{\sigma}$, coverage errors will not vanish completely.

Let $\gamma^*_{1-\alpha}$ denote a $(1-\alpha)$-quantile of S^*, then the resulting left-sided CI is:

$$\begin{aligned} \mathbb{P}\left(S^* \leq \gamma^*_{1-\alpha}\right) &= 1-\alpha \\ \Rightarrow \mathbb{P}\left(T - \hat{\sigma}\cdot\gamma^*_{1-\alpha} \leq g(\theta)\right) &\approx 1-\alpha \end{aligned} \quad (2.12)$$

Remark 3:
An important difference between BQM and Bootstrap-t is their switched usage of quantiles: BQM uses $G^{*,-1}(\alpha)$ (α-quantile of resampled T^* values), but Bootstrap-t uses $\gamma^*_{1-\alpha}$ ($(1-\alpha)$-quantile of resampled S^* values). The upcoming hybrid bootstrap will shad more light on the version of Bootstrap-t.

2.3.4 Bootstrap Principle and Hybrid Bootstrap

Bootstrap Principle

Assume that an unknown quantity $g(\theta)$ is estimated by $g(\hat{\theta})$, where $F = F_\theta$ is the sampling distribution. In the bootstrap community it is preferred to write $F_0 = F$, $F_1 = \hat{F}$ and $g(\theta) = T(F_0)$, $g(\hat{\theta}) = T(F_1)$. Many statistical problems now take the form:

$$\text{Population equation (solve for } t\text{): } \mathbb{E}(f_t(F_0, F_1) \mid F_0) = 0 \quad (2.13)$$

for a suitable functional f_t. The solution t_0 of equation (2.13) is the quantity of interest, e.g.:

[3] e.g. using the jackknife, see [15], or a bootstrap algorithm (double bootstrap)

- **Calculation of the true bias**
  Set $f_t(F_0, F_1) = T(F_1) - T(F_0) + t$, then the solution t_0 of the population equation is the bias of $T(F_1)$.

- **Exact $(1-\alpha)$-CI for $T(F_0)$**
  Set
  $$f_t(F_0, F_1) = \mathbb{1}_{[T(F_1)-t,\, T(F_1)+t]}(T(F_0)) - (1-\alpha), \tag{2.14}$$
  then $[T(F_1) - t_0,\ T(F_1) + t_0]$ will be an exact $(1-\alpha)$-CI for $T(F_0)$.

Solving (2.13) requires F_0, hence the **bootstrap principle**[4] substitutes (F_0, F_1) by (F_1, F_2) to estimate t_0:

$$\text{Sample equation: } \mathbb{E}(f_t(F_1, F_2) \mid F_1) = 0 \tag{2.15}$$

F_2 denotes an estimate of F_1, when resampled data $X^* \sim F_1$ are used.

Hybrid Bootstrap

Applying the bootstrap principle to quantile-CI gives the sample equation:

$$\mathbb{E}\Big(\mathbb{1}_{[T(F_2)-\hat{t},\, T(F_2)+\hat{t}]}(T(F_1)) \mid F_1\Big) = 1-\alpha$$

Inserting the solution $\hat{t}_0$ to $[T(F_1) - t_0,\ T(F_1) + t_0]$ gives the CI $[T(F_1) - \hat{t}_0,\ T(F_1) + \hat{t}_0]$. For left-sided CI, the functional

$$f_t(F_0, F_1) = \mathbb{1}_{[S(F_1)-t,\infty)}(S(F_0)) - (1-\alpha),$$

is needed, giving:

$$\mathbb{E}\Big(\mathbb{1}_{[T(F_2)-\hat{t}_0,\infty)}(T(F_1)) \mid F_1\Big) = \mathbb{P}\big(T(F_2) - \hat{t}_0 \le T(F_1) \mid F_1\big) = 1-\alpha$$

It follows that $\hat{t}_0$ has to be the $(1-\alpha)$-quantile of $T(F_2)-T(F_1)$. Notice, that BQM would use the α-quantile of $T(F_2)$ to construct left-sided CI, contrary to the bootstrap principle. A direct consequence of this observation is the **hybrid bootstrap** (see [16]), applying BQM to $n^l(T(F_1) - T(F_0))$ (for some constant l). In practice hybrid bootstrap is preferred to bypass the calculation of $\hat{\sigma}$ (via jackknife or nested bootstrap) for each bootstrap sample.

The resulting CI is given by:

$$\mathbb{P}\big(T - \hat{\sigma}\gamma^*_{1-\alpha} \le g(\theta)\big) \approx 1-\alpha \tag{2.16}$$

Here $\gamma^*_{1-\alpha}$ is the $(1-\alpha)$-quantile of $T^* - T$. If G_Δ denotes the distribution of $T - g(\theta)$ and G the distribution of T, then it holds:

$$\begin{aligned} 1-\alpha &= G(\gamma^*_{1-\alpha}) = \mathbb{P}\big(T - g(\theta) \le \gamma^*_{1-\alpha}\big) = G_\Delta(g(\theta) + \gamma^*_{1-\alpha}) \\ \Rightarrow G^{*,-1}(1-\alpha) &= T + G^{*,-1}_\Delta(1-\alpha) = T + \gamma^*_{1-\alpha} \end{aligned}$$

Hence, equation (2.16) is taking the form:

$$\boxed{\mathbb{P}\big(2T - G^{*,-1}(1-\alpha) \le g(\theta)\big) \approx 1-\alpha}$$

[4] As formulated in [17]

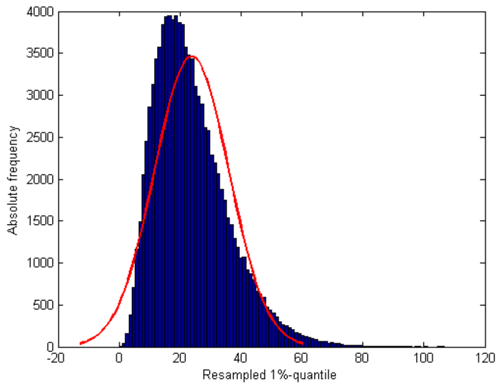

Figure 2.1: Histogram of 100,000 bootstrap 1%-quantiles.

2.3.5 Application to Survival Data

Negative CIs

Lifetimes are nonnegative, but for positively skewed G hybrid Bootstrap (as well as Bootstrap-t) CI reaching into the negative numbers might appear:

Example 1:
10 data were randomly drawn from a $\mathcal{W}(100; 2)$-distribution:

34, 45, 71, 102, 114, 126, 127, 141, 169, 215

The Weibull models MLE are $\hat{\beta} = 129.25$, $\hat{\gamma} = 2,35$. 100,000 bootstrap samples were generated, each time the 1%-quantile was estimated.

Figure 2.1 shows a histogram of the resampled quantiles $\hat{q}^*_{0.01}$. A left-sided 5%-CI can be read of as $[8.03,\ \infty)$ (=BQM), a right sided 95%-CI as $[0,\ 46.78]$. The hybrid Bootstrap ($l = 1$) uses the MLE $\hat{q}_{0.01} = 18.22$ and $G^{*,-1}(0.95) = 46.78$ to give the CIs left side:

$$\hat{q}_{p,\alpha} = 2 \cdot 18.22 - 46.78 = -10.34$$

.

Logarithmic Transformations

The problem of negative CI can be fixed using logarithmic transformations: Transform the sample to $\underline{Y} = \underline{\ln X}$, get a quantile-CI $\hat{q}^{\ln}_{p,\alpha}$ and transform it back to $\hat{q}_{p,\alpha} = \exp \hat{q}^{\ln}_{p,\alpha}$. Step by step this means:

1. Relation between quantiles of X and $\ln X$:

$$\mathbb{P}(X \leq s) = \mathbb{P}(\ln X \leq \ln s) \Rightarrow q^{\ln}_p = \ln q_p$$

2. Consequence for quantile estimates: $\hat{q}^{\ln}_p = \ln \hat{q}_p$

3. Analogously for G^* quantiles: $G^{*,-1}_{\ln}(1-\alpha) = \ln G^{*,-1}(1-\alpha)$

4. Write down the logarithmic quantile (hybrid bootstrap):

$$\hat{q}_{p,\alpha}^{\ln} = 2 \cdot \hat{q}_p^{\ln} - G_{\ln}^{*,-1}(1-\alpha)$$

5. Transform $\hat{q}_{p,\alpha}^{\ln}$ back:

$$\begin{aligned}\hat{q}_{p,\alpha} &= e^{\hat{q}_{p,\alpha}^{\ln}} = e^{2\cdot\hat{q}_p^{\ln} - G_{\ln}^{*,-1}(1-\alpha)} \\ &= \boxed{\hat{q}_p \cdot \frac{\hat{q}_p}{G^{*,-1}(1-\alpha)}}\end{aligned}$$

Example 2 (Example 1 continued):
Using the logarithmic approach for the hybrid Bootstrap, we get the following 95%-CI for 1%-quantile:

$$\hat{q}_{0.01,0.05} = 18.22 \cdot \frac{18.22}{46.78} = 7.10$$

Bias-Corrected Logarithmic Hybrid Bootstrap

As a final possible refinement, bias-corrected BQM might be used to get the CI for $T^* - T$. This means using $G^{*,-1}\left(\Phi\left(2z_0 + \Phi^{-1}(1-\alpha)\right)\right)$ instead of $G^{*,-1}(1-\alpha)$. Remember that $z_0 = \Phi^{-1}(G^*(T))$.

Example 3 (Example 1 continued):
The 95%-quantile 46.78 of G^* has to be replaced by a different quantile. First of all, $G^*(T) = 0.379$, i.e. 37.9% of the resampled 1%-quantile were less or equal than $\hat{q}_{0.01} = 18.22$. It follows that $z_0 = \Phi^{-1}(0.379) = -0.3081$. Hence we have to calculate the $\Phi(2z_0 + \Phi^{-1}(1-\alpha) = \Phi(2\cdot(-0.3081) + 1.6449) = 0.8482$-quantile of G^*. Finally $G^{*,-1}(0.8482) = 35.61$, giving the quantile-CI:

$$\hat{q}_{0.01,0.05} = 18.22 \cdot \frac{18.22}{35.61} = 9.32$$

In practical applications not only one quantile is calculated, but a variety of values $p \in [0,1]$ is used for reasons of visualisation. This might lead to some undesirabilities as shown in the next example.

Example 4:
Assume that the following censored sample was observed:

$$80,288,\ 69,241,\ 100,000^+,\ 53,514,\ 59,009,\ 100,000^+$$

A 95%-CI for several $p \in [0,1]$ is calculated using hybrid bootstrap and its bias corrected version. The results are displayed in a **probability net**. This means, that the CDF is plotted in coordinate systems with axes transformations:

$$x' = \ln x,\ y' = \ln(-\ln(1-y))\ \text{(inverse of standard Gumbel)}$$

Due to this transformation, every Weibull CDF $y = F(x)$, $F = \mathcal{W}(\beta;\gamma)$ will appear as a straight line $y' = \gamma x' - \gamma \ln \beta$. Figure 2.2 shows, that for small p the CI of both methods are similar. For $p > 0.2$ the bias-corrected hybrid bootstrap starts

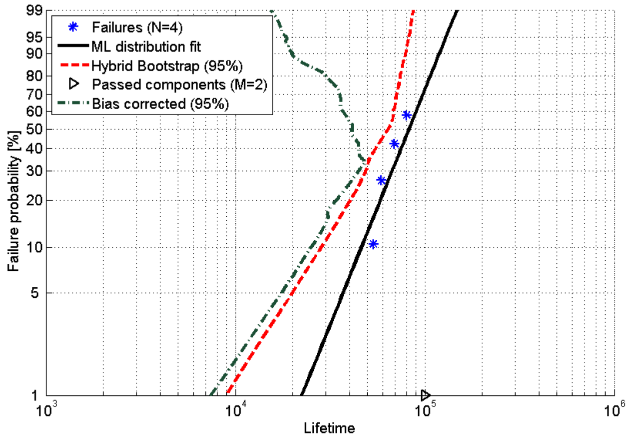

Figure 2.2: Confidence lines of the hybrid bootstrap and its bias-corrected version.

showing kinks, roughly between two neighbouring data points. The influence of the bias correction starts to get very large for growing p, making the mapping $\hat{q}_{p,\alpha} \mapsto p$ not even injective. E.g. the lifetime 40,000 is the 30%- as well as the 75% quantile CI.

To get a consistent visualisation, the bias-corrected version should not be used for censored data.

2.4 Comparison of the Quantile CI Estimators

To compare the logarithmic delta method with the bias-corrected logarithmic hybrid Bootstrap a small simulation study was done. A Weibull sample, with $\beta = 80,000$, $\gamma \in \{1.8, 2.5\}$, of size $N = 10$ was chosen. All data where censored[5] at $L = 100,000$, leading to a censoring probability of 77.56% resp. 82.57 (depending on γ). To calculate coverage errors of left-sided 95%-CI for $q_{0.01}$ (true value 6,211 resp. 12,704) Monte Carlo simulation is used[6].

Table 2.1 shows the empirical confidences of the methods for $N = 10$. While the hybrid Bootstrap's coverage error is less than 5%, the delta method has one of about 10%. Event further, censoring has almost no effect on the bootstraps's coverage error, while for the delta method it has a positive effect. For this small sample size, resampling methods work much better than asymptotic methods. For reasons of completeness, table 2.2 shows coverage errors for $N = 20$.

While the delta method improves slightly to coverage errors of 5%, the hybrid Bootstraps coverage error decreases to less than 2%. This time censoring has suprisingly a positive effect on the delta method, while the bootstrap is almost not touched. **The overall result of our simulations fits the general believe, that resampling methods do well for small samples.** In

[5]values in brackets belong to $L = \infty$, i.e. no censoring
[6]Simulation size depending on method: $M_{\text{Delta}} = 10,000$, $M_{\text{Boot}} = 1,000$

γ	Method	Emp. confidence	CI (95%)
1.8	Delta	86.85%	[86.17%,87.51%]
		(83.20%)	(82.54%,84.02%])
	hybrid Boot	89.10%	[87.00%,90.96%]
		(90.60%)	([88.62%,92.34%])
2.5	Delta	85.91%	[85.21%,86.59%]
		(83.09%)	([82.34%,83.82%])
	hybrid Boot	91.20%	[89.27%,92.88%]
		(89.70%)	([87.65%,91.51%])

Table 2.1: Coverage errors for $N = 10$ and $\gamma = 1.8$ ($\gamma = 2.5$).

γ	Method	Emp. confidence	CI (95%)
1.8	Delta	89.47%	[88.85%,90.07%]
		(86.57%)	([85.89%,87.23%])
	hybrid Boot	94.30%	[92.68%,95.65%]
		(93.90%)	([92.23%,95.30%)
2.5	Delta	89.33%	[88.71%,89.93%]
		(85.69%)	([84.99%,86.37%])
	hybrid Boot	93.70%	[92.01%,95.13%]
		(94.70%)	([93.12%,96.01%])

Table 2.2: Coverage errors for $N = 20$.

practice further reduction of the coverage error is achieved by fixing the shape parameter, as is done in the next section.

2.5 Hypothesis Testing using Quantile Estimates

Section 1.5 formulated a quantile hypothesis $H_0 : F(q_0) \geq p_0$, to be rejected. Reliability estimation compares estimates quantile-CI $\hat{q}_{p_0,\alpha}$ to the design life. Emerging coverage errors will lead to a test having wrong significance.

2.5.1 Monte Carlo Corrected Delta Method

Reliability estimation technique (RET) uses test statistic $\hat{q}_{p_0,\alpha}$ and critical value q_0. Since quantile-CI methods are not exact, the nominal significance α is not achieved. Monte Carlo methods may be used to decrease the coverage error.

General concept: Let $T \mid H_0 \sim F_0$ be a test statistic. The true critical values of T are given by quantiles of F_0, which may be approximated by the empirical CDF of simulated F_0 realisations. Algorithmic cheap T, like the delta method, are desirable. In anticipation to section 3.2 we restrict ourselves to **one-parameter distribution models** $F \in \{F_\theta \mid \theta \in \Theta \subseteq \mathbb{R}\}$ (typically achieved by fixing the distributions shape parameter to worst-case-values). This claim is needed to determine θ uniquely by H_0 for lognormal and Weibull distributions.

Remark 4 (Delta method for one parameter distributions):
Fixing the shape parameter in equation (2.7) allows to use normal- instead of

t-quantiles. The asymptotic covariance matrix reduces to a real number. For details see appendix B.

MC corrected delta method: Design a test with significance α for a sample of size N (censored at L)

1. Determine the unique θ_0, such that $F_{\theta_0}(q_0) = p_0$.
2. Draw M samples $\underline{X}_i^*$ of size N from F_{θ_0}.
3. Censor each sample at L (failure censoring also possible).
4. Apply the delta method to each resample to get $\hat{q}^{*,i}_{p_0,\alpha}$.
5. Calculate the α-quantile q_{crit} of $\hat{q}^{*,i}_{p_0,\alpha}$.

2.5.2 Design of Experiments

Remark 1 already addressed the problem of completely censored samples, making the calculation of quantile CI impossible. Since such samples belong to unexceptional large lifetimes, we decide for H_1. Consequently the experimental design must be in a way, such that completely censored samples are unlikely under H_0. If the maximal test duration is given by L, then the H_0-probability for completely censored samples is:

$$\mathbb{P}_{H_0}(\delta_1 = \cdots = \delta_N = 0) = (1 - F_{\theta_0}(L))^N \overset{!}{\leq} \alpha \tag{2.17}$$

We refer to this as the **significance (un)equation**. Chapter 3 will take deeper look at this equation. For the moment we just claim, that N and L are chosen adequate.

Remark 5 (Minimal sample sizes):
Most statistical test only require minimal sample sizes if the error of second kind is to be bounded. In our case, censoring leads to discontinuous distribution functions, requiring minimal sample sizes already to ensure a small error of first kind.

2.5.3 Discontinuities and Resulting Power Functions

The delta methods critical value q_0 was replaced by a Monte Carlo calculated quantile q_{crit}. This quantile is a compromise between usual and degenerated quantile-CI estimation, as is shown in the next example.

Example 5:
We want to test for $H_1 : F(10,000) < 0.01$. The lifetimes are modeled to be Weibull with $\gamma \leq 1.8$. If $N = 8$ prototypes are available, using equation (2.17), a test duration $L \geq 9.47\dot{q}_0$ is needed ($\alpha = 0.01$).

If $H_0 : F(10,000) \geq 0.01$ is true, then $\beta \leq \beta_0 = 128,795$. Figure 2.3 shows the quantile estimates of 2,000 $\mathcal{W}(\beta_0;\gamma)$-samples. The main proportion of results lies in the area 3,000-7,000, and some values isolated at 10,000, belonging to completely censored samples (single observations censoring probability: 53.04%, sample probability $0.5304^8 = 0.0062 = 0.62\%$).

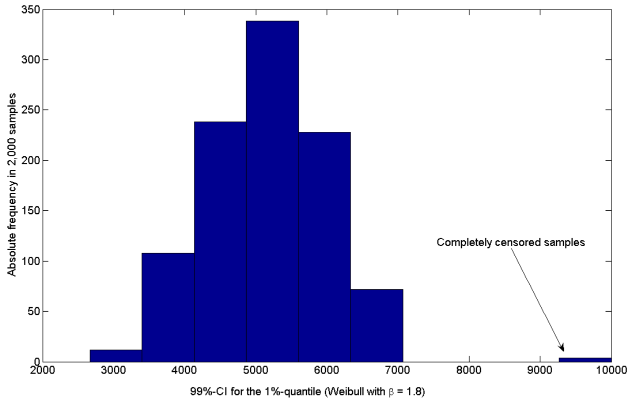

Figure 2.3: Histogram of 20,000 simulated quantile-CI under H_0 for $L = 10q_0$.

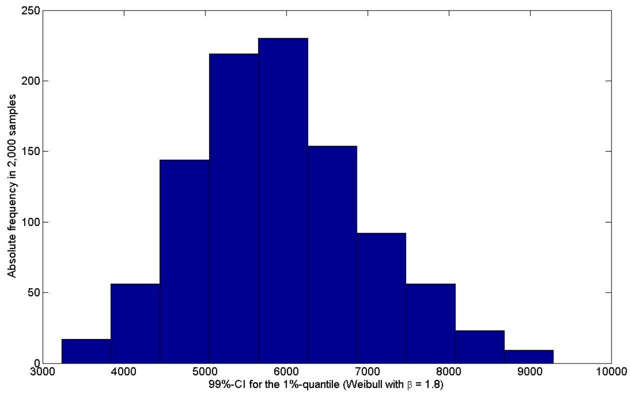

Figure 2.4: Histogram of 20,000 simulated quantile-CI under H_0 for $L = 15q_0$.

The reason for this separation (discontinuity) is of course the small sample size resp. the short test duration. E.g. larger N would lower the probability of completely censored samples on one hand (isolated values appear with less probability), and reduce the mean squared error of the non-degenerated estimates on the other hand (main proportion of quantile estimates would move toward $q_0 = 10,000$). Figure 2.4 shows this effect for an increased test duration.

The delta methods coverage error would hence lead to an empirical significance of 0.62% (sample censoring probability) instead of the desired 1% (this means only completely censored samples lead to H_1). Due to histogram 2.3 we get a corrected critical value of $q_{\text{crit}} = 6,692$ (99%-quantile of the histogram). In case of failures, the quantile-CI will exceed this value only with probability $1\% - 0.62\% = 0.38\%$, giving the corrected delta method the correct significance.

We are now interested in the influence of correcting critical values on the tests power function.

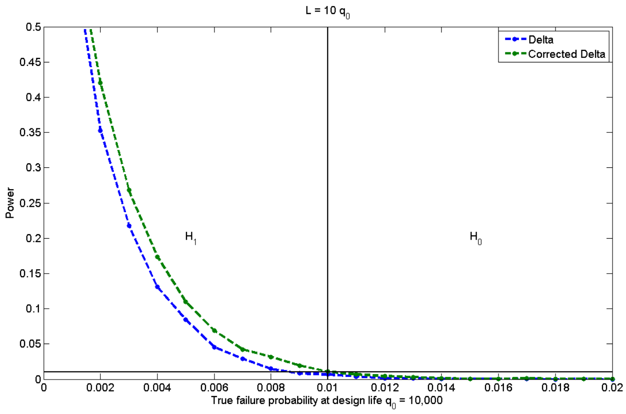

Figure 2.5: Delta method's power functions for $L = 10q_0$.

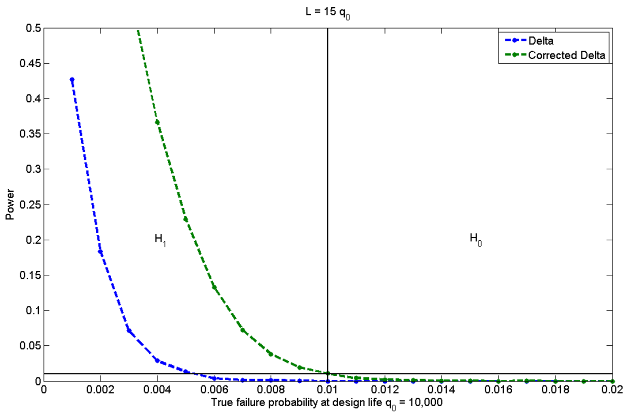

Figure 2.6: Delta method's power functions for $L = 15q_0$.

Example 6 (Example 5 continued):
At each position $F(10,000) = p_i$, $p_i = 0.001 + i \cdot 0.001$, $i \in \{0, \dots, 19\}$ the power was calculated using 10,000 Monte Carlo samples[7]. Figure 2.5 shows, that the empirical significance of about 0.62% is fixed by the correction.

In the case $L = 15q_0$, leading to a censoring probability of 26.83%, it is very unlikely to get a completely censored sample: $0.2683^8 = 2.6851 \cdot 10^{-5}$. However the mse of the quantile estimates is still large, making it unlikely to get values beyond $q_0 = 10,000$. As a consequence the power is very poor and using $q_{\text{crit}} = 8,638$ has a large impact on the power (see figure 2.6)

Further power functions can be found in chapter 8, where the corrected delta method will be compared to counting methods.

[7] see also section 8.1

2.6 Summary

Using quantile estimates for hypothesis testing has to deal with the problem of completely censored samples, leading to degenerated quantile estimates. Granting homologation in these cases requires:

$$\mathbb{P}_{H_0}(\delta_1 = \cdots = \delta_N = 0) = (1 - F_{\theta_0}(L))^N \overset{!}{\leq} \alpha \tag{2.18}$$

To get a unique left hand side, we have to assume one-parameter distribution models (e.g. by fixing shape to worst-case values). The numbers N, L have to be chosen accordingly. As soon as failures are observed, the delta quantile-CI is formally defined, but has a considerable coverage error. To fix the coverage error of the resulting test, a new critical value has to be calculated using Monte Carlo methods.

The effect of this correction is large if many failures are observed (figure 2.6), since otherwise most homologations are due to completely censored samples (figure 2.5) (i.e. no usage of the critical value), leading to discontinuities in the test statistics distribution. **Overall, using reliability estimation was not convincing for small samples sizes**.

Chapter 3

Success Runs for Reliability Demonstration

Success runs try to demonstrate the claimed reliability $H_1 : F(q_0) < p_0$ by comparing the postulated failure quota p_0 at time q_0 with the observed one. If a one parameter distribution model F_θ, $\theta \in \mathbb{R}$ is given, one can shift the hypothesis to time $L \geq q_0$. The effect of hypothesis transformation on the power function will be studied, showing the model's drawback.

3.1 Nonparametric Distribution Models

3.1.1 Test Statistic and Acceptance Region

Success runs simply count the number of units surviving the test duration. Since the exact lifetime is not considered, one speaks of an **attribute life test**.

Definition 8 (Number of passed components):
Let $T_1, \ldots, T_N \overset{\text{iid}}{\sim} F$ then the number of passed components (w.r.t. q_0) is defined as:

$$S_{q_0} = |\{i \mid T_i \geq q_0\}| \in \mathbb{N}$$

An immediate consequence is the binomial distribution of S_{q_0}:

$$\mathbb{P}(S_{q_0} = k) = \binom{N}{k}(1 - F(q_0))^k F(q_0)^{N-k} \tag{3.1}$$

giving this approach the name **binomial sampling**

Remark 6:
With a view to the upcoming CUS sampling, S_{q_0} can be interpreted as follows: Each passed component (i.e. $T_i > q_0$) is valued with one point, each failure (i.e. $T_i \leq q_0$) with zero points. The test statistic S_{q_0} then is the total number of points achieved with the sample.

The validity of H_0 means, that for k near to N, the probability

$$\mathbb{P}(S_{q_0} = k) = \binom{N}{k}(1 - F(q_0))^k F(q_0)^{N-k}$$

can be bounded from above by

$$\mathbb{P}_{H_0}(S_{q_0} = k) \leq \binom{N}{k}(1-p_0)^k p_0^{N-k}$$

Large S_{q_0} indicate the correctness of the alternative $H_1 : F(q_0) < p_0$, i.e. choose the critical value S_{crit} such that the **significance equation** holds:

$$\mathbb{P}_{H_0}(S_{q_0} \geq S_{\text{crit}}) \leq \alpha \tag{3.2}$$

The frequently used **success run method** uses the critical value $S_{\text{crit}} = N$ (no failures before time L), giving the significance equation:

$$\mathbb{P}_{H_0}(S_{q_0} \geq N) = (1-p_0)^N \leq \alpha \tag{3.3}$$

Using $S_{\text{crit}} < N$ is sometimes called **generalised success run** or **success run with failures**.

3.1.2 Minimal Sample Sizes

Choosing smaller critical values than $S_{crit} = N$ increases the minimal sample size in 3.2. Therefore the sample size required to fulfil is given by equation (3.3) as:

$$N_{\min} = \left\lceil \frac{\ln \alpha}{\ln(1-p_0)} \right\rceil \tag{3.4}$$

and no failure may occur

Since p_0 is very small $\ln(1-p_0)$ may be approximated by the Taylor polynomial $(1-p_0) - 1 = -p_0$. Therefore the minimal sample size is of order:

$$N_{\min} \approx \frac{-\ln \alpha}{p_0} \tag{3.5}$$

E.g. $\alpha = 0.01$ implies that $N_{\min} \approx 4.6 \cdot p_0^{-1}$. Therefore the verification of p_0 in **ppm (parts per million)** would require testing several million prototypes.

Example 7:
For $p_0 = 0.01$, $q_0 = 10{,}000$, $\alpha = 0.05$ one finds:

S_{crit}	N	N-1	N-2	N-3	N-4
$N_{\min}$	299	473	628	773	913

This means: If 299 units were tested without failure, homologation is achieved. If one failure occurred before time q_0, then an additional number of 473-299=174 units have to be observed without failure before time q_0. These sample sizes are much too high for practical applications.

3.1.3 Success Runs for Small Lots

Assume that N units are tested from a lot of size N_0 (in section 3.1.1 N_0 was not specified, but assumed to be large). If D_0 of the N_0 units are defective, the number D of units tested defective follows a hypergeometric distribution:

$$\mathbb{P}(D = k) = \frac{\binom{D_0}{k}\binom{N_0-D_0}{N-k}}{\binom{N_0}{N}}$$

The lot is accepted if at most D_{crit} failures are observed:

$$\mathbb{P}(D \leq D_{\text{crit}}) = \sum_{k=0}^{D_{\text{crit}}} \frac{\binom{D_0}{k}\binom{N_0-D_0}{N-k}}{\binom{N_0}{N}} \overset{!}{\leq} \alpha$$

In case of $D_{\text{crit}} = 0$ it holds:

$$\frac{(N_0 - N)\cdots(N_0 - N - D_0 + 1)}{N_0 \cdots (N_0 - D_0 + 1)} \leq \alpha$$

allowing for numerical calculation of the minimal sample size.

For large N_0 a binomial approximation can be used (giving success runs):

$$\mathbb{P}(D \leq D_{\text{crit}}) \approx \sum_{k=0}^{D_{\text{crit}}} \binom{N}{k} p^k (1-p)^{N-k},\ p = \frac{D_0}{N_0}$$

Alternatively via Poisson distributions:

$$\mathbb{P}(D \leq D_{\text{crit}}) \approx \sum_{k=0}^{D_{\text{crit}}} \frac{N^k p^k}{k!} e^{-Np},\ p = \frac{D_0}{N_0}$$

Binomial approximations require $N \leq 0.1 \cdot N_0$, Poisson ones additionally $p < 0.1$ and normal ones large sample sizes to give $Np(1-p) > 9$ (see [9]). The number of produced components is large compared to any tested sample, making us use the binomial approximation.

3.2 One-Parameter Distribution Models

Consider parametric distribution models, e.g. Weibull or lognormal[1]. From experience (or tables like in [1] or [2]) the shape parameter is held at some worst case value, describing the productions scatter. The unknown scale parameter θ is determined uniquely under the hypothesis boundary $\partial H_0 : F_\theta(q_0) = p_0$. The hypothesis might hence be shifted to the censoring time L. Counting the components passing time L requires $p_L = F_\theta(L)$ to derive critical values for S_L.

3.2.1 Hypothesis Transformation

Consider any one-parameter lifetime distribution with the property that the mapping $\theta \mapsto F_\theta(q)$ is bijective for every q. Weibull and lognormal distributions with fixed γ resp. σ^2 are of such type. The initial hypothesis $H_0 : F_\theta(q_0) \geq p_0$

[1]As introduced in appendix A

determines θ_0 uniquely at its boundary $F_\theta(q_0) = p_0$. Consequently, a shifted hypothesis can be formulated: $H_0 : F_\theta(L) \geq p_L$, where $p_L = \mathbb{P}(T \leq L \mid \theta = \theta_0)$. Due to monotonicity of F and $q_0 < L$ it will hold that $p_L > p_0$. This property will allow for small minimal sample sizes in the next section.

3.2.2 Test Statistic and Acceptance Region

The reason for the enormous sample sizes in (3.4) was the verification of a small failure probability p_0 (see equation 3.5). Shifting the hypothesis to increased test durations L helps lowering N.

Lemma 1 *If the hypothesis $H_0 : F_\theta(q_0) \geq p_0$ is transformed to $H_0' : F_\theta(L) \geq p_L$ with $L > q_0$, then for the resulting minimal sample sizes (see equation (3.4)) it holds:*

$$N_{\min}(q_0) \geq N_{\min}(L)$$

Proof

$$(1 - p_0) \geq (1 - p_L) \Rightarrow \alpha \geq (1 - p_0)^N \geq (1 - p_L)^N$$

$$\Rightarrow N_{\min}(p_0) = \left\lceil \frac{\ln \alpha}{\ln(1 - p_0)} \right\rceil \geq N_{\min}(p_L) = \left\lceil \frac{\ln \alpha}{\ln(1 - p_L)} \right\rceil$$

qed

Example 8 (Lognormal distributions):
The lifetime of the component is modeled lognormally distributed $\mathcal{LN}\left(\mu; \sigma^2\right)$ with unknown scale μ and $\sigma \leq 0.6$. It follows under H_0:

$$\begin{aligned} p_0 &= \Phi\left(\frac{\ln q_0 - \mu_0}{\sqrt{\sigma^2}}\right) \\ \Leftrightarrow \mu_0 &= \ln q_0 - \sigma \cdot \Phi^{-1}(p_0) \\ \Leftrightarrow p_L &= \Phi\left(\frac{1}{\sigma} \ln \frac{L}{q_0} + \Phi^{-1}(p_0)\right) \end{aligned}$$

Hence, the failure quota at time L does only depend on the ratio $\lambda = \frac{L}{q_0}$ of test duration and design life. Table 3.1 shows the impact of this transformation on the sample size.

Example 9 (Weibull distributions):
The lifetime of the component is modeled Weibull-distributed $\mathcal{W}(\beta; \gamma)$ with unknown scale β and $\gamma \geq 2$. Under H_0:

$$\begin{aligned} p_0 &= 1 - e^{-\left(\frac{q_0}{\beta_0}\right)^\gamma} \\ \Leftrightarrow \beta_0 &= q_0 \cdot (-\ln(1 - p_0))^{-\frac{1}{\gamma}} \\ \Leftrightarrow p_L &= 1 - (1 - p_0)^{\left(\frac{L}{q_0}\right)^\gamma} \end{aligned}$$

Again, p_L depends only on the ratio $\lambda = \frac{L}{q_0}$. The sample size $N_{\min}(L)$ can be calculated as:

$$N_{\min}(L) = \left\lceil \frac{\ln \alpha}{\ln(1 - p_L)} \right\rceil = \left\lceil \lambda^{-\gamma} \frac{\ln \alpha}{\ln(1 - p_0)} \right\rceil$$

λ	$N - S_{\text{crit}}(L)$	$N_{\min}(L)$
5	0 (success run)	5
	1	8
	2	10
10	0 (success run)	2
	1	4
	2	5
20	0 (success run)	1
	1	2
	2	4

Table 3.1: Acceptance region and sample size for $p_0 = 0.01$, $q_0 = 10,000$, $\alpha = 0.01$ for a lognormal distribution with $\sigma^2 \leq 0.36$

λ	$N - S_{\text{crit}}(L)$	$N_{\min}(L)$
5	0 (success run)	19
	1	27
	2	35
10	0 (success run)	5
	1	8
	2	10
20	0 (success run)	2
	1	3
	2	4

Table 3.2: Acceptance region and sample size for $p_0 = 0.01$, $q_0 = 10,000$, $\alpha = 0.01$ for a Weibull distribution with $\gamma \geq 2$

3.3 Power of the Test

Transforming the hypothesis to large L secures the applicability of the success run approach by decreasing the sample size. Looking at the resulting power functions reveals a major drawback. For reasons of simplification only $S_{\text{crit}} = N$ is considered, but the results transfer to the general case.

Definition 9 (Power of a test):
Let A be the acceptance region for H_1 for a statistical test with test statistic T. While the error of first kind (significance) is given by $\mathbb{P}_{H_0}(T \in A)$, the error of second kind is given by $\mathbb{P}_{H_1}(T \notin A)$. If $F = F_\theta$, $\theta \in \Theta$ and $H_0 : \theta \in \Theta_0 \subseteq \Theta$, $H_1 : \theta \in \Theta \setminus \Theta_0$, then the ***power of the test*** *at position θ is defined by:*

$$g(\theta) = \mathbb{P}(T \in A \mid F = F_\theta)$$

In this sense, the power describes the opposite event to an error of second kind.

3.3.1 Calculation of the Power Function

The hypothesis H_0 is rejected, if and only if all units pass time L. Let $g_L(p)$ denote the probability of this event, given $F_\theta(q_0) = p$ (i.e. $H_0 : p \geq p_0$, $H_1 : p < p_0$), then:

$$g_L(p) = (1 - F_{\theta_p}(L))^{N_{\min}(L)}$$

While an ideal test would imply

$$g(p) = \begin{cases} 0, & p \geq p_0 \\ 1, & p < p_0 \end{cases}$$

the integer nature of N implies that in general no equality will hold in

$$g_L(p_0) = (1 - p_L)^N \leq \alpha \tag{3.6}$$

as will be explained in the next subsection.

3.3.2 Properties of the Power Function

The shortfall in equation (3.6) will increase for increasing L:

- If $1 - F_{\theta_p}(L) \approx 1$, a large N is needed for $(1 - F_{\theta_0}(L))^N \leq \alpha$. Simultaneously the difference between $(1 - F_{\theta_0}(L))^i$ and $(1 - F_{\theta_0}(L))^{i+1}$ is small, making the sequence slowly decreasing w.r.t i.
- If $1 - F_{\theta_p}(L) \ll 1$, a small N is needed for $(1 - F_{\theta_0}(L))^N \leq \alpha$. Now the difference between $(1 - F_{\theta_0}(L))^i$ and $(1 - F_{\theta_0}(L))^{i+1}$ is larger, making the sequence decrease faster.

Consequently for the power functions and sample sizes of $L_1 < L_2$:

$$N_{\min}(L_1) \geq N_{\min}(L_2) \text{ but } \alpha \geq g_{L_1}(p_0) \geq g_{L_2}(p_0)$$

The impact on the rest of the power function can be seen in fig. 3.1: The conservative behavior increases with L, i.e. feasible components will enter production less likely.

Definition 10 (Consumers and producers risk):
In connection to the homologation of components, the errors of first and second kind are renamed:

- ***Consumers risk:*** *Probability of granting homologation for unreliable designs (error of first kind).*
- ***Producers risk:*** *Probability of refusing homologation for reliable designs (error of second kind).*

Using these new risk notions we can summarise: **For fixed α, every feasible experimental design (N, L) bounds the consumers risk by α, but the producers risks increases for decreasing N.**

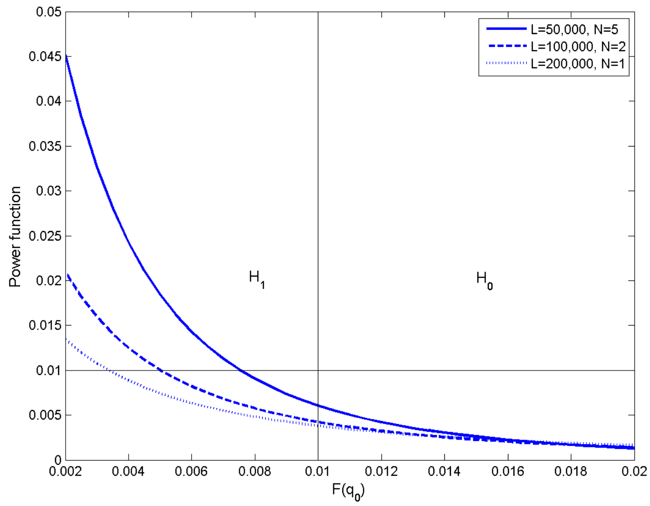

Figure 3.1: Power function for different L with $p_0 = 0.01, q_0 = 10,000, \alpha = 0.01$, for a lognormal distribution with $\sigma^2 = 0.36$. Sample sizes due to table 3.1.

3.4 The Suitability of Lognormal and Weibull Models for Success Runs

3.4.1 The Hazard of Lognormal Distributions

In appendix A.1.2 the hazard of a lognormal distribution is shown to be non-monotonic with a maximum. Only left of this maximum wearout failures are the modeled failure mechanism (see figure A.1). If lognormal models are used to derive success runs, one has to take care, that the test duration L is left of the maximum (or even left of the point of inflection).

For reasons of simplicity we study which values of p_L might appear and how these values depend on the distributions parameter. First consider pdf and cdf of a lognormal distribution:

$$\begin{aligned} f(t) &= \frac{1}{\sigma t}\phi\left(\frac{\ln t - \mu}{\sigma}\right) \\ F(t) &= \Phi\left(\frac{\ln t - \mu}{\sigma}\right) \end{aligned}$$

The hazard (as introduced in appendix A) is defined to be

$$h(t) = \frac{f(t)}{1 - F(t)}$$

Lemma 2 *Let x_0 be the maximum of the hazard h of a lognormal model, then $F(x_0)$ does not depend on μ.*

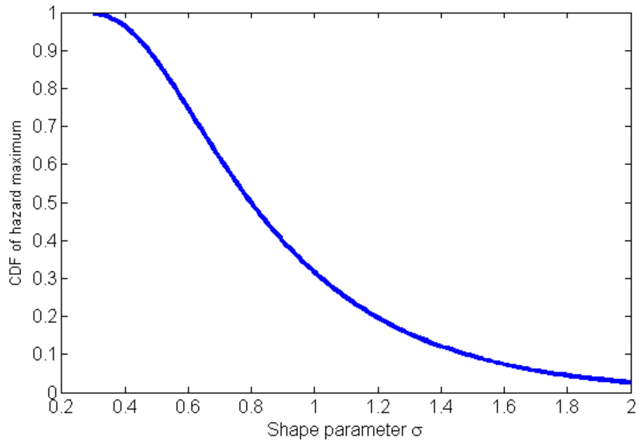

Figure 3.2: CDF of the lognormal hazards maximum.

Proof *The PDF, CDF and hazard of a lognormal distribution are:*

$$h'(t) = \frac{f'(t)(1-F(t)) + f^2(t)}{(1-F(t))^2}$$

Let $t_0 = e^{\mu+\sigma z_0}$ *denote the maximum of* h*, then it holds* $h'(t_0) = 0$*. Therefore:*

$$\begin{aligned} 0 &= f'(t)(1-F(t)) + f^2(t) \\ &= \left(\frac{-1}{\sigma t^2}\phi\left(\frac{\ln t - \mu}{\sigma}\right) + \frac{1}{\sigma t}\phi'\left(\frac{\ln t - \mu}{\sigma}\right)\cdot\frac{1}{\sigma t}\right)(1-F(t)) + f^2(t) \\ &= \left(\frac{-1}{\sigma t^2}\phi(z_0) + \frac{1}{\sigma t}\phi'(z_0)\cdot\frac{1}{\sigma t}\right)(1-\Phi(z_0)) + \left(\frac{1}{\sigma t}\right)^2\phi^2(z_0) \\ \Leftrightarrow 0 &= (\phi'(z_0) - \sigma\phi(z_0))\,(1-\Phi(z_0) + \phi^2(z_0) \end{aligned}$$

It follows, that z_0 *does only depend on* σ*. Finally, the CDF is given by* $F(t_0) = \Phi(z_0)$*, which gives the claim.*

qed

An explicit calculation of the maximum is not possible, since there is no closed formula for the lognormal CDF. Numerical maximisation of the hazard was done to visualise $F(x_0)$ as a function of σ in figure 3.2.

It can be seen, that e.g. $\sigma = 0.6$ does allow for $p_L \leq 0.7455$. In example 8 there was a design $\lambda = 10$, $N = 2$, requiring $1 - p_L \leq \alpha^{1/N} = 0.1$, i.e. $p_L \geq 0.9$. This design is therefore doubtful, since its test duration lies clearly beyond the hazards maximum. Due to figure 3.2 a smaller σ would allow for such a p_L, e.g. $\sigma \approx 0.4$.

3.4.2 Weibull Analysis

Consider again example 9 where the formula

$$p_L = 1 - (1-p_0)^{\left(\frac{L}{q_0}\right)^\gamma}$$

appeared. Using $R_L = 1 - p_L$, $R_0 = 1 - p_0$ gives $R_0 = R_0^{\lambda^\gamma}$ (compare: appendix A.2.3). The significance equation of the success run does take the simple form:

$$R_0^{N \cdot \lambda^\gamma} \leq \alpha$$

There is no comparable formula for lognormal distributions, distinguishing Weibull distributions for success runs. Due to this convincing relation between N, L, α the success run is strongly connected to Weibull models, giving the method also the name **Weibull analysis**.

3.5 Summary: Binomial models

The number N, needed to test the hypothesis $H_0 : F(q_0) = p_0$, can be calculated using binomial models for the number of passed components S_L:

$$N_{\min} = \left\lceil \frac{\ln \alpha}{\ln(1 - p_L)} \right\rceil$$

One-parameter distribution models reduce N by accounting for the censoring time L. Success runs ($S_{\text{crit}} = N$) offer the smallest sample sizes. Whenever N gets too small, the power function decreases and the success run ends up being conservative (increased producers risk). The reason for this conservative behavior is the integer nature of S_{crit} and S_L.

Using lognormal models causes the problem, that large L might leave the area, where wearout failures are modeled (depending only on σ, which is required to be small). On the contrary, Weibull distributions fit very well to the success runs update due to their quantile relations.

Chapter 4

Randomisation of Success Runs

Randomisation is a classical approach to address the problem of integer valued critical values. First the general concept is described, secondly it will be applied to success runs. After one is getting improved power functions the problem of repeatability is discussed.

4.1 Randomisation for General Tests

Let T be a test statistic and A be an acceptance region for the alternative H_1. To keep the significance α, A is chosen such that $\mathbb{P}_{H_0}(T \in A) \leq \alpha$. If T is integer valued, then typically $\mathbb{P}_{H_0}(T \in A) < \alpha$ and $\mathbb{P}_{H_0}(T \in B) > \alpha$ for any $B \supset A$. So, the significance can not be kept exactly. Now, randomisation chooses randomly between the acceptance regions A and B, where B is chosen as small as possible.

Let $H \in \{0, 1\}$ with $\mathbb{P}(H = 1) = \phi_B$, independent of T. For $H = 1$ the set B is used, for $H = 0$ the set A is used as acceptance region. Then the probability of rejecting H_0 is:

$$\begin{aligned} \alpha &\stackrel{!}{=} (1 - \phi_B)\mathbb{P}_{H_0}(T \in A) + \phi_B \mathbb{P}_{H_0}(T \in B \setminus A) \\ \Rightarrow \phi_B &= \frac{\alpha - \mathbb{P}_{H_0}(T \in A)}{\mathbb{P}_{H_0}(T \in B \setminus A)} \end{aligned} \tag{4.1}$$

From equation 4.1 one derives the power function of the randomised test:

$$g_{\text{Rand}} = (1 - \phi_B)g_A + \phi_B g_B \tag{4.2}$$

where g_A, g_B are the power functions from the tests with acceptance regions A, B. By construction $g_{\text{Rand}}(p_0) = \alpha$, but also the rest of the power function is improved:

Lemma 3 *On the whole domain of the power functions it holds that:*

$$g_A \leq g_{\text{Rand}} \leq g_B$$

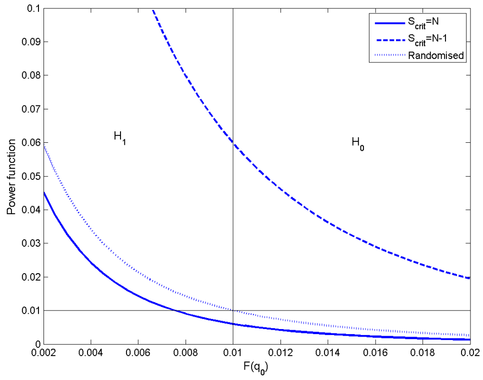

Figure 4.1: Power of success runs after randomisation.

Proof *Since $A \subset B$ it follows immediately that $\mathbb{P}(S_L \in A) \leq \mathbb{P}(S_L \in B)$ showing $g_A \leq g_B$. Now $\phi_B \in [0,1]$, which means, that g_{Rand} is by equation (4.2) contained in the convex hull of g_A and g_B.*

qed

4.2 Application for Success Runs

For generalised success runs A is the event that at most $N - S_{\text{crit}}$ failures occur and B is the event that at most $N - S_{\text{crit}} + 1$ failures occur. For the case of success runs:

$$\begin{aligned} \mathbb{P}_{H_0}(S_L = N) &= (1-p_L)^N \\ \mathbb{P}_{H_0}(S_L = N-1) &= Np_L(1-p_L)^{N-1} \\ \Rightarrow \phi_B &= \frac{\alpha - (1-p_L)^N}{Np_L(1-p_L)^{N-1}} \end{aligned} \tag{4.3}$$

Of course this formula for ϕ_B can be extended to any other critical value $S_{\text{crit}} < N$ in complete analogy.

Example 10:
For a lognormal distribution with $\sigma^2 = 0.36$ and $p_0 = 0.01$, $q_0 = 10,000$, $\alpha = 0.01$, $L = 5q_0$. With $p_L = 0.9349$ and equation (4.3) one is deriving $\phi_B = 0.0735$. This means in 92,65% of the the evaluations one claims no failure for among the $N = 5$ tested units, in the remaining 7,35% of possible cases one failure is allowed to achieve homologation.

Figure 4.1 shows how the powers of the initial success runs with $S_{\text{crit}} = N$ and $S_{\text{crit}} = N - 1$ are combined to a test with $g(p_0) = \alpha$.

4.3 Acceptance of Randomised Tests

From a mathematical point of view the randomisation approach is correct, but it is unlikely that it will be adopted in practice due to legal aspects. The verification of safety is based on the result of a coin toss, letting components enter production despite failures where observed. Every evaluation of the test results might therefore lead to different decisions about homologation. Therefore we present a different approach to fix the success runs drawback in the next chapter.

Chapter 5

Partially-Passed Component Counting

Success runs without hypothesis transformation require large sample sizes, success runs with hypothesis transformation behave conservatively. As already mentioned, the reason for this behavior lies in the discrete test statistic and/or critical value. In this chapter a generalisation of the success run method is introduced, allowing to count every failure as a partial success to improve the methods power.

5.1 Motivation

The current test statistic of the success run can be written as:

$$S_L = |\{T_i \mid T_i \geq L\}| = \sum_{i=1}^{N} \mathbb{1}_{[L,\infty)}(T_i)$$

Components with $T_i \geq L$ contribute 1 to S_L, otherwise 0, no matter how close T_i is to L. The latter information is dismissed completely

Definition 11 (PPC count):
*Let $H_0 : F_{\theta_0}(q_0) \geq p_0$ be a hypothesis in an one parameter model $\theta_0 \in \Theta \subseteq \mathbb{R}$, then the **partially-passed component (PPC)** count test statistic B_L is defined as:*

$$B_L = \sum_{i=1}^{N} F_{\theta_0}(T_i^+) \tag{5.1}$$

where T_i^+ denotes the right-censored (at L) observation of T_i.

Using F_{θ_0} is based on the fact, that $F_{\theta_0}(T_i)$ would lead to an uniform distribution under H_0. Using T_i^+ instead of T_i will lead to the densitiy stated in Theorem 3.

Remark 7:
The statistic $\frac{1}{F_{\theta_0}(L)} B_L$ can be interpreted as follows[1]:

- If $T_i = 0$ then $\frac{1}{F_{\theta_0}(L)} F_{\theta_0}(0) = 0$ and no points are achieved.
- If $T_i = L$ then $\frac{1}{F_{\theta_0}(L)} F_{\theta_0}(L) = 1$.
- Results $0 < T_i^+ < L$ are scoring in $(0, 1)$, depending on the relation between T_i^+ and L in Terms of F_{θ_0}.

In particular $\frac{1}{F_{\theta_0}(L)} B_L = S_L$ for success runs.

5.2 The Distribution of B_L

The **simulation lemma** states that $F(T)$ follows a uniform distribution if $T \sim F$. In case of B_L we need a version of the simulation lemma for sums of $F(T^+)$.

Theorem 3
Let $T_1, \ldots, T_N \overset{\text{iid}}{\sim} F_\theta$, $\theta \in \Theta \subseteq \mathbb{R}$, $H_0 : F_{\theta_0}(q_0) = p_0$, $\bar{F} = 1 - F$ *and* $\theta \mapsto F_\theta(q)$ *be bijective for every* q, *further* $T_1^+, \ldots, T_N^+$ *be the right censored versions of* $T_1, \ldots, T_N$ *at time* L, *then* $B_L(T_1^+, \ldots, T_N^+)$ *has under* H_0 *the following PDF:*

$$\begin{aligned} g_N^+(u) &= \sum_{i=0}^{N-1} \binom{N}{i} F_{\theta_0}(L)^{N-1-i} \bar{F}_{\theta_0}(L)^i g_{N-i}\left(\frac{u}{F_{\theta_0}(L)} - i\right) \\ &\quad + \bar{F}_{\theta_0}(L)^N \delta_{N F_{\theta_0}(L)}(u),\ u \in [0, N F_{\theta_0}(L)] \end{aligned} \tag{5.2}$$

here δ *is a delta distribution:*

$$\int_A \delta_x(s) ds = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

and g_k *denotes the PDF of the* k*-fold convoluted uniform distribution on* $[0, 1]$, *i.e. the density of* $U_1 + \cdots U_k$ *where* U_i *is uniform distributed on* $[0, 1]$.

Proof ***Induction on*** N, *see Appendix C.*

qed

Remark 8:
For Weibull distributions, Theorem 3 will turn out to be a special case of theorem 7 in section 6.1, were sums of beta distributions appear, since $F \neq F_0$.

In the special case of $F_{\theta_0}(L) = 1$, $g_N^+(u) = g_N(u)$ is simply the convolution of N uniform distributions on $[0, 1]$.

[1] See also remark 6

5.3 Properties of the Distribution of B_L

The approach of using the PPC count will hence be the **censored uniform sum (CUS)** model, in analogy to the binomial models for S_L.

Definition 12:
A random variable with PDF

$$\begin{aligned} g_N^+(u) &= \sum_{i=0}^{N-1} \binom{N}{i} \rho^{N-1-i}(1-\rho)^i g_{N-i}\left(\frac{u}{\rho} - i\right) \\ &\quad +(1-\rho)^N \delta_{N\rho}(u),\ u \in [0, N\rho(L)] \end{aligned} \tag{5.3}$$

(compare Theorem 3 with $\rho = F_{\theta_0}(L)$) will be called $\mathcal{CUS}(N, \rho)$ distributed.

Theorem 4
Let $X \sim \mathcal{CUS}(N, \rho)$, $Y \sim \mathcal{CUS}(M, \rho)$ be independent, then:

$$\mathbb{E}(X) = N\frac{\rho(2-\rho)}{2},\ \mathbb{V}\text{ar}(X) = N\frac{\rho^3(4-3\rho)}{12} \tag{5.4}$$

$$X + Y \sim \mathcal{CUS}(N + M, \rho) \tag{5.5}$$

Proof *From the proof of theorem 3 it is known, that*

$$F(s) = s + \mathbb{1}_{[\rho,1]}(s) \cdot (1-\rho),\ s \geq 0,$$

is the distribution fucntion of a $\mathcal{CUS}(1, \rho)$ variable. It follows:

$$\begin{aligned} \mathbb{E}(\mathcal{CUS}(1,\rho)) &= \int_0^\rho s \cdot (1 + \delta_\rho(s) \cdot (1-\rho))\, ds \\ &= \int_0^\rho s ds + (1-\rho) \int_0^\rho s\delta_\rho(s) ds \\ &= \frac{\rho^2}{2} + (1-\rho)\rho \end{aligned}$$

$$\begin{aligned} \mathbb{V}\text{ar}(\mathcal{CUS}(1,\rho)) &= \int_0^\rho s^2 \cdot (1 + \delta_\rho(s) \cdot (1-\rho))\, ds - \left(\frac{\rho(2-\rho)}{2}\right)^2 \\ &= \int_0^\rho s^2 ds + (1-\rho) \int_0^\rho s^2\delta_\rho(s) ds - \left(\frac{\rho(2-\rho)}{2}\right)^2 \\ &= \frac{\rho^3}{3} + (1-\rho)\rho^2 - \frac{\rho^2(2-\rho)^2}{4} \end{aligned}$$

Now, the property $X + Y \sim \mathcal{CUS}(N + M, \rho)$ and the formulas for expectation and variance follow immediately from the fact that $\mathcal{CUS}(N, \rho)$ is the distribution of the sum of N i.i.d. $\mathcal{CUS}(1, \rho)$-variables.

qed

5.4 Calculation of Critical Values

5.4.1 Exact calculation

The alternatives H_1 acceptance region for $\frac{1}{F_{\theta_0}(L)} B_L$ is chosen in the form

$$\left[\frac{m_{\text{crit}}}{F_{\theta_0}(L)},\ N\right]$$

where m_{crit} is determined by:

$$\int_{m_{\text{crit}}}^{NF_{\theta_0}(L)} g_N^+(u)du = \alpha \tag{5.6}$$

Since the PDF is known from theorem 3 one is getting with $k = \left\lfloor \frac{m_{\text{crit}}}{F_{\theta_0}(L)} \right\rfloor$:

$$\begin{aligned}
\alpha &= \int_{m_{\text{krit}}}^{NF_{\theta_0}(L)} g_N^+(u)du \\
&= \sum_{i=0}^{N-1} \binom{N}{i} F_{\theta_0}(L)^{N-i} \overline{F}_{\theta_0}(L)^i \int_{\frac{m_{\text{krit}}}{F_{\theta_0}(L)}-i}^{N-i} g_{N-i}(s)ds + \overline{F}_{\theta_0}(L)^N \\
&= \underbrace{\sum_{i=k+1}^{N} \binom{N}{i} F_{\theta_0}(L)^{N-i} \overline{F}_{\theta_0}(L)^i}_{\alpha_R(k)} \\
&\quad + \sum_{i=0}^{k} \binom{N}{i} F_{\theta_0}(L)^{N-i} \overline{F}_{\theta_0}(L)^i \int_{\frac{m_{\text{krit}}}{F_{\theta_0}(L)}-i}^{N-i} g_{N-i}(s)ds
\end{aligned}$$

In the special case $F_{\theta_0}(L) = 1$:

$$\alpha = \int_{m_{\text{krit}}}^{N} g_N(s)ds$$

For numerical calculations a minimal k is chosen with $\alpha_R(k) \leq \alpha$ first (i.e. k is the critical value of a success run with failures), secondly a solution of

$$\alpha - \alpha_R(k) = \sum_{i=0}^{k} \binom{N}{i} F_{\theta_0}(L)^{N-i} \overline{F}_{\theta_0}(L)^i \int_{\frac{m_{\text{crit}}}{F_{\theta_0}(L)}-i}^{N-i} \Delta_{N-i}(s)ds$$

is searched in the interval $[kF_{\theta_0}(L); (k+1)F_{\theta_0}(L)]$ (e.g. via bisection). **The equations left hand side is recognised to be the success runs coverage error**. The necessary densities Δ_{N-i} can be calculated explicitly for small $N-i$, for large $N-i$ be approximated by normal distributions, compare section 5.4.2.

Corollary 1 *The minimal sample size for CUS sampling is given by:*

$$(1 - F_{\theta_0}(q_0))^N \leq \alpha \tag{5.7}$$

Proof *Assume equation (5.7) will not hold, then $NF_{\theta_0}(L)$ will have point mass greater α. Since this point is also the minimal possible acceptance region for the alternative, the desired significance can not be reached.*

qed

This means, that the minimal sample size is the same for CUS and binomial samples[2]. However, the PPC count does not show conservative behavior at point p_0. **Switching from binomial to CUS sampling improves power without increasing the sample size**.

Remark 9:
The problem with the quantile estimates significance equation was, that they are not defined for completely censored samples. Success runs and binomial models are well defined for this case, but their CDFs do have point masses that should be below the significance. Hence, the reasons for required minimal sample sizes are slightly different, leading to the same equation.

Example 11:
To reject the hypothesis $H_0 : F_{\theta_0}(10,000) \geq 0.01$ with significance $\alpha = 0.01$ one can test $N = 5$ units. Their lifetime is supposed to be Weibull distributed with $\gamma \geq 1.8$. In the first step, the necessary test duration L has to be calculated. From the significance equation for Weibull distributions we have:

$$\begin{aligned} \left(1 - \left(1 - e^{-\left(\frac{L}{\beta_0}\right)^{\gamma_0}}\right)\right)^N &\leq \alpha \\ e^{-\left(\frac{L}{\beta_0}\right)^{\gamma_0}} &\leq \sqrt[N]{\alpha} \\ \Rightarrow L &\geq \beta_0 \cdot \left(-\frac{\ln \alpha}{N}\right)^{\frac{1}{\gamma_0}} \end{aligned}$$

It follows for the scale parameter β_0 under H_0:

$$\begin{aligned} 1 - e^{-\left(\frac{q_0}{\beta_0}\right)^{\gamma_0}} &\geq p_0 \\ \Leftrightarrow \beta_0 &\geq \frac{q_0}{(-\ln(1-p_0))^{\frac{1}{\gamma_0}}} \end{aligned}$$

$$\Rightarrow L \geq q_0 \cdot \left(\frac{\ln \alpha}{N \cdot \ln(1-p_0)}\right)^{\frac{1}{\gamma_0}}$$

For the chosen values one has $L \geq 123,042$ ($\beta_0 \geq 128,795$), which might be rounded towards $L = 125,000$. In this setup, success runs allow no failure:

$$p_L = F_{\theta_0}(125,000) = 1 - e^{-\left(\frac{125,000}{}\right)^{1.8}} = 0.6123$$

$$\Rightarrow (1 - p_L)^4 > \alpha,\ (1 - p_L)^5 \leq \alpha$$

[2] As well as quantile estimates, see section 2.5.2

CUS sampling uses $(1-\alpha)$-quantile m_{crit} of a $\mathcal{CUS}(5, 0.6123)$ distribution, which is $m_{\text{crit}} = 3.05$. The sample then has to achieve $\frac{3.05}{0.6123} = 4.98$ points out of 5 possible. This could happen for example, if the first four tested units pass $L = 125,000$ and the fifth unit achieves 0.98 points, which means:

$$F_{\theta_0}(X_5) \geq 0.98 \cdot 0.6123 \Rightarrow X_5 \geq 122,699$$

5.4.2 Approximation with Normal Distributions

Since B_L is a sum iid random variables (having finite mean and variance according to theorem 4), the central limit theorem with $\rho = F_{\theta_0}(L)$ applies:

Theorem 5
Let $B_1, \ldots, B_N$ be $\mathcal{CUS}(N, \rho)$-distributed, then $\frac{1}{N}B_N$ converges in distribution to a normal distribution:

$$\sqrt{N}\frac{\frac{1}{N}B - \frac{\rho(2-\rho)}{2}}{\sqrt{\frac{\rho^3(4-3\rho)}{12}}} \stackrel{N\to\infty}{\longrightarrow} \mathcal{N}(0;1)$$

Proof *By theorem 4, B_N can be decomposed into N iid summands $X_1, \ldots, X_N$ with distribution $\mathcal{CUS}(1, \rho)$. Each X_i has finite mean and variance:*

$$\mathbb{E}(X_i) = \mu = \frac{\rho(2-\rho)}{2}, \ \mathbb{V}\mathrm{ar}(X_i) = \sigma^2 = \frac{\rho^3(4-3\rho)}{12}$$

By the central limit theorem the random variable:

$$\frac{B_N - N\mu}{\sigma\sqrt{N}} = \sqrt{N}\frac{\frac{1}{N}B - \frac{\rho(2-\rho)}{2}}{\sqrt{\frac{\rho^3(4-3\rho)}{12}}}$$

converges in distribution to a normal distribution.

qed

For large N one can consequently try to use the approximation:

$$B \sim \mathcal{N}\left(N\frac{F_{\theta_0}(L)(2-F_{\theta_0}(L))}{2}; N\frac{F_{\theta_0}(L)^3(4-3F_{\theta_0}(L))}{12}\right)$$

The quality of this approach depends strongly on ρ, which is deciding if the single summand distributions are symmetric. From figures 5.1, 5.2 and 5.3 one can see, that a small ρ requires a larger N to get a good normal approximation. As a rule of thumb the approximation with normal distributions is reasonable if $N\rho \geq 4$.

5.5 PPC for Other Censoring Patterns

Up to now, the described new counting method is heavily connected to time censored sampling. If CUS sampling shall be applied to a censoring pattern like failure censored sampling or sudden death sampling (to be defined in the next subsection), the whole method has to be adapted. Success runs are not used for such censoring patterns, since the resulting samples do typically consist of some failures (information a success run would ignore). PPC counting will allow to introduce counting methods for these patterns without loosing information.

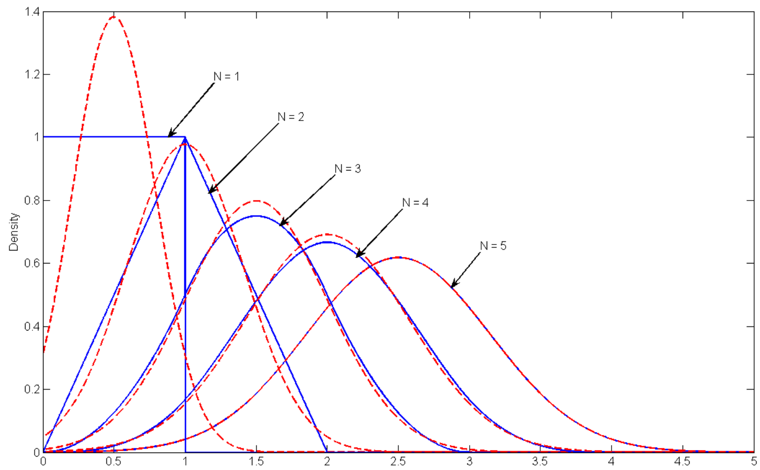

Figure 5.1: Approximation with normal distributions for $\rho = 1$

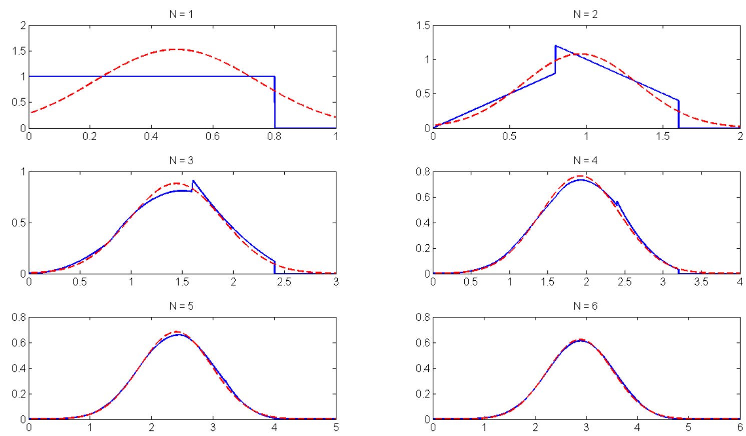

Figure 5.2: Approximation with normal distributions for $\rho = 0.8$

5.5.1 Failure Censored Sampling

In failure censored sampling one always waits for the first r failures to happen. All other experiments still running are suspended. Therefore, one observes failures $T_{(1)}, \ldots, T_{(r)}$ and $N-r$ suspended items at time $T_{(r)}$. The PPC count for failure censored samples is therefore:

$$B_r = \sum_{i=1}^{r} F_{\theta_0}\left(T_{(i)}\right) + (N-r) \cdot F_{\theta_0}\left(T_{(r)}\right) \tag{5.8}$$

Different to time censored sampling B_r will not reach the maximum value N. Typically, a failure censored sample is not even evaluated with counting methods at all. Nevertheless, CUS sampling can be adapted as soon as a critical value is available.

Since F_{θ_0} defines a monotonic increasing transformation, $F_{\theta_0}\left(T_{(i)}\right) \sim U_{(i)}$ where $T \sim F_{\theta_0}$ and $U_{(r)}$ is the r-th order statistic out of N uniform variables. **The distribution of B_L (under H_0) therefore only depends on N and r.**

Lemma 4 *The r-th order statistic of N Uniform distributions is Beta distributed with parameters* $\mathrm{Beta}(r; N+1-r)$.

Proof *A uniform distribution has PDF $f(x) = 1$ and CDF $F(x) = x$. In general, the PDF of the r-th order statistic is given by:*

$$f_{(r)}(x) = \frac{N!}{(r-1)!(N-r)!} F^{r-1}(x) \cdot f(x) \cdot (1-F(x))^{N-r}$$

(See e.g. [9]). Plugin in f, F gives the claim.

qed

Unfortunately, since $T_{(1)}, \ldots, T_{(r)}$ are not independent, the distribution of B_r can not be calculated by convoluting the PDFs from lemma 4. Especially, B_r is not CUS distributed, but an **o**rdered **u**niform **s**um (OUS(N, r)). Monte Carlo simulation might be used to calculate the critical values for B_L.

Example 12:
Consider again the values of example 11. Instead of suspending every component surviving time $L = 125,000$, $r = 3$ failures are claimed. To get the distribution of B_r one million realisation of $U_{(1)}, U_{(2)}, U_{(3)}$ where simulated. Using lemma 4 a simulated value for B_r is obtained as:

$$B_r = U_{(1)} + U_{(2)} + 3U_{(3)}$$

Figure 5.4 shows the histogram of the one million simulated values. The critical value for B_r is given by the 99%-quantile of simulated values, which is 3.8008.

Remark 10:
B_r is a linear combination of order statistics (L-estimate, see [14]). There is a huge amount of literature for this topic, mainly deriving asymptotic results. Since r is small, asymptotic theory can not be applied.

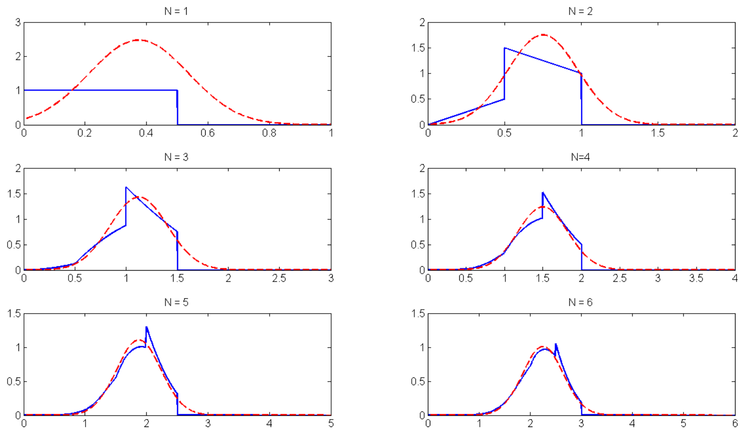

Figure 5.3: Approximation with normal distributions for $\rho = 0.5$

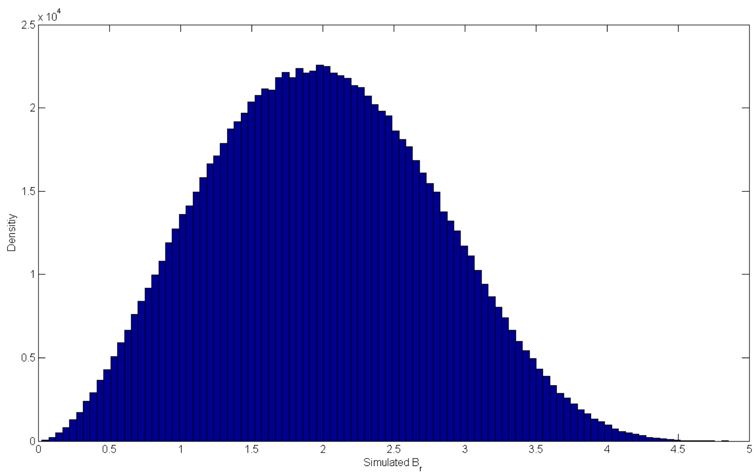

Figure 5.4: PPC count for failure censoring with $N = 5$, $r = 3$.

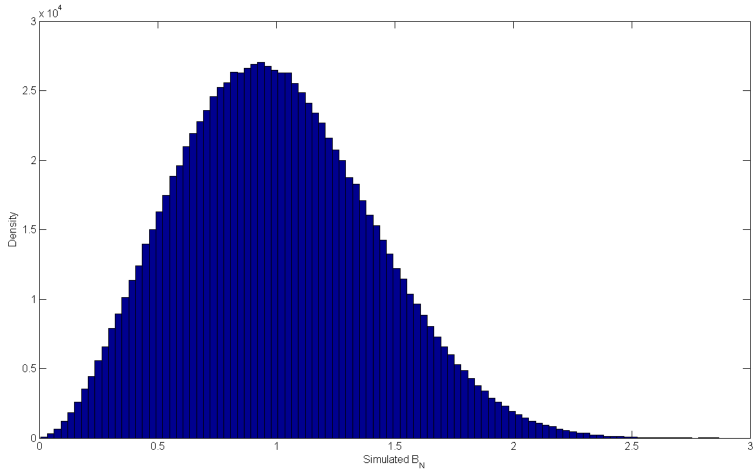

Figure 5.5: PPC count for sudden death censoring with $N = 3$ pairs.

5.5.2 Sudden Death Sampling

In sudden death sampling units are tested pairwise (e.g. wheel carriers in a system test). Let $T_{i,1}, T_{i,2}$ denote the random lifetimes of the i-th pair (out of N pairs). When the first component fails, both are suspended, giving PPC count:

$$B_N = 2 \cdot \sum_{i=1}^{N} \min\left(F_{\theta_0}\left(T_{i,1}\right), F_{\theta_0}\left(T_{i,2}\right)\right)$$

The single summand B_N follows an ordered uniform sum distribution

$$\text{OUS}(2,1) = \text{Beta}(1;2)$$

having PDF $2(1-x)$. All N pairs are independent, making $\frac{1}{2}B_N$ the convolution of N Beta(1; 2)-variables. This time, the distribution of the PPC count B_N might be called **OUS21-sum**, depending only on the number of pairs N. One simple way of getting critical values is again given by Monte Carlo simulation.

Example 13:
If 3 pairs of components are tested, the PPC count has the distribution of figure 5.5 (calculated by 1.000.000 Monte Carlo simulations).

Figure 5.4 shows the histogram of the one million simulated values. The critical value for B_N is given by the 99%-quantile of simulated values, which is 2.0178.

5.6 Summary

CUS sampling was introduces as a generalisation of binomial sampling, accounting for the exact lifetimes:

$$\frac{1}{p_L} B_L = \frac{1}{p_L} \sum_{i=1}^{N} F_{\theta_0}(T_i^+)$$

Critical values may be calculated numerically, allowing for a test that has the correct significance. Further properties of the power function will be discussed in chapter 8. The test statistics distribution unter H_0 does depend on the sample size N and the censoring probability $F_{\theta_0}(L)$. The PPC count can be considered as a direct generalisation of the success run. Both counting methods require the same minimal sample size:

$$N_{\min} = \left\lceil \frac{\ln \alpha}{\ln(1 - p_L)} \right\rceil$$

For other censoring patterns PPC counting is also available. **Due to its usage of all information, PPC counting can be used for all three censoring patterns, and one does not need to switch between success runs** (time censoring with small L or failure censoring with small r) **and quantile estimates** (time censoring with large L or failure censoring with large r).

Chapter 6

Maximum Likelihood for CUS Sampling

Up to now we only considered the distribution of $\frac{1}{p_L} B_L$ under H_0 to derive critical values for reliability demonstration (RDT). It is left to derive the likelihood of B_L under other values of the unknown distribution parameters to get a reliability estimation (RET) method from censored uniform sum (CUS) sampling. This serves as a basis for next chapters about Bayesian methods.

6.1 The Likelihood of CUS Sampling

To develop the theory of CUS sampling for the case $T \not\sim F_0$, Weibull distributions have to be assumed. Weibulls special quantile structure (see section A.2.3) leads to convolutions of beta distributions.

Lemma 5 *Assume that $T \sim \mathcal{W}(\beta; \gamma)$ and γ to be the fixed nuisance parameter. If β_0 denotes the unknown parameter under hypothesis H_0, then it holds:*

$$F_{\beta_0,\gamma}(T) \sim \text{Beta}\left(1; \left(\frac{\beta_0}{\beta}\right)^{\gamma}\right)$$

Proof *The CDF of $F_{\beta_0,\gamma}(T)$ can immediately be derived as:*

$$\begin{aligned} \mathbb{P}(F_{\beta_0,\gamma}(T) \leq s) &= \mathbb{P}\left(T \leq F^{-1}_{\beta_0,\gamma}(s)\right) = 1 - e^{-\left(\frac{\beta_0(-\ln(1-s))^{\frac{1}{\gamma}}}{\beta}\right)^{\gamma}} \\ &= 1 - (1-s)^{\left(\frac{\beta_0}{\beta}\right)^{\gamma}} \end{aligned}$$

The PDF is given by the derivative:

$$\frac{d}{ds}\mathbb{P}(F_{\beta_0,\gamma}(T) \leq s) = \left(\frac{\beta_0}{\beta}\right)^{\gamma} \cdot (1-s)^{\left(\frac{\beta_0}{\beta}\right)^{\gamma}-1}$$

which is recognised to be the claimed beta density.

qed

Remark 11:
The case $\beta = \beta_0$ gives $F_0(T) \sim \text{Beta}(1;1)$ which is a uniform distribution. This shows that Theorem 3 is a special case of Theorem 7 below.

Theorem 6
Let T^+ be the censored version of $T \sim F = \mathcal{W}(\beta;\gamma)$ at time L and $F_0 = \mathcal{W}(\beta_0;\gamma)$ be the distribution under H_0, then $B_L = F_0(T^+)$ has the following PDF:

$$g_1^+(u) = g_1(u) \cdot \mathbb{1}_{[0,F_0(L)]}(u) + (1 - F(L))\delta_{F_0(L)}(u) \tag{6.1}$$

where g_1 denotes the PDF of a $\text{Beta}\left(1; \left(\frac{\beta_0}{\beta}\right)^\gamma\right)$*-variable (see lemma 5).*

Proof *Given $T \leq L$, the random variables T, T^+ can not be distinguished. In case of $T > L$ the CDF G_1^+ has to collect all mass of G_1 in the single point F_0. It follows:*

$$G_1^+(s) = G_1(\min(s, F_0)) + \mathbb{1}_{[F_0(L),1]}(s) \cdot \bar{F}$$

Hence the PDF is:

$$g_1^+(s) = g_1(\min(s, F_0)) \cdot \mathbb{1}_{[0,F_0]}(s) + \bar{F} \cdot \delta_{F_0}(s), \; s \in [0, F_0(L)]$$

qed

Remark 12:
When deriving the likelihood from (6.1) one has to account for the dependency of $F(L)$ on the sampling distributions parameter β additionally to the dependency of g_1 on β. The likelihood is now given by a product

$$L(R \mid T_1^+, \ldots, T_N^+) = \prod_{i=1}^{N} g_1^+\left(F_0(T_i^+)\right),$$

where R is the component's probability to survive time L.

A formal expression for the distribution of B_L for general N may be derived, but requires numerics.

Theorem 7
Let $T_1, \ldots, T_N \overset{\text{iid}}{\sim} \mathcal{W}(\beta;\gamma)$, $H_0 : F_{\beta_0}(q_0) = p_0$, further $T_1^+, \ldots, T_N^+$ be the right censored versions of $T_1, \ldots, T_N$ at time L, then $B_L(T_1^+, \ldots, T_N^+)$ has the following PDF:

$$\begin{aligned} g_N(u) &= \sum_{i=0}^{N-1} \binom{N}{i} F_{\theta_0}(L)^{N-1-i} \bar{F}_{\theta_0}(L)^i g_{N-i}^+\left(\frac{u}{F_{\theta_0}(L)} - i\right) \\ &\quad + \bar{F}_{\theta_0}(L)^N \delta_{NF_{\theta_0}(L)}(u), \; u \in [0, NF_{\theta_0}(L)] \end{aligned} \tag{6.2}$$

here δ is a delta distribution:

$$\int_A \delta_x(s)ds = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

and g_k^+ is denotes the PDF of k independent convoluted $\text{Beta}\left(1; \left(\frac{\beta_0}{\beta}\right)^\gamma\right)$*-variables.*

Proof *Simple imitation of the proof in section C, where Theorem 6 serves as induction start.*

6.2 Maximum Likelihood Estimation of CUS Sampling

Using Theorem 7 the log-likelihood for the reliability R (probability of surviving time L) is given by:

$$\begin{aligned} l(R \mid \underline{T}^+) &= \sum_{i=1}^{N} \ln \left(g_1\left(F_0(T_i^+)\right) \cdot \mathbb{1}_{[0,F_0(L)]}\left(F_0(T_i^+)\right) + (1-F(L))\delta_{F_0(L)}(F_0(T_i^+)) \right) \\ &= \sum_{\delta_i=1} \ln g_1\left(F_0(T_i^+)\right) + \sum_{\delta_i=0} \ln\left((1-F(L))\delta_{F_0(L)}(F_0(L))\right) \\ &= \sum_{\delta_i=1} \ln B(R) + \sum_{\delta_i=1} (B(R)-1)\cdot \ln(1-F_0(T_i^+)) \\ &\quad + \ln(1-F(L)) \cdot \sum_{\delta_i=0} 1 \\ &= N_u \ln B(R) + (B(R)-1) \sum_{\delta_i=1} \ln(1-F_0(T_i^+)) + N_c \lambda^\gamma \ln R \end{aligned}$$

where $N_u = \sum \delta_i$, $N_c = N - N_u$, $B(R) = \left(\frac{\beta_0}{\beta(R)}\right)^\gamma$ and $F(L) = F_{\beta(R)}(L)$ depends on R (see Theorem 6). The derivatives of l require the ones of $B(R)$:

$$\begin{aligned} B(R) &= -\left(\frac{\beta_0}{q_0}\right)^\gamma \cdot \ln R \\ B'(R) &= -\frac{\left(\frac{\beta_0}{q_0}\right)^\gamma}{R} \\ B''(R) &= \frac{\left(\frac{\beta_0}{q_0}\right)^\gamma}{R^2} \end{aligned}$$

Additionally the derivatives of the function $R \mapsto 1 - F_{\beta(R)}(L)$ are needed:

$$\begin{aligned} 1 - F_{\beta(R)}(L) &= R^{\lambda^\gamma} \\ \frac{\partial}{\partial R}(1 - F_{\beta(R)}(L))) &= \lambda^\gamma R^{\lambda^\gamma - 1} \\ \frac{\partial^2}{\partial R^2}(1 - F_{\beta(R)}(L))) &= \lambda^\gamma (1-\lambda^\gamma) R^{\lambda^\gamma - 2} \end{aligned}$$

It follows:

$$\frac{\partial}{\partial R} l(R \mid \underline{T}^+) = \frac{N_u}{R \cdot \ln R} - \frac{\left(\frac{\beta_0}{q_0}\right)^\gamma}{R} \cdot \sum_{\delta_i=1} \ln(1-F_0(T_i^+)) + \frac{N_c \lambda^\gamma}{R}$$

Hence, the MLE is:

$$\frac{N_u}{\ln \hat{R}} + N_c \lambda^\gamma = \left(\frac{\beta_0}{q_0}\right)^\gamma \cdot \sum_{\delta_i=1} \ln(1-F_0(T_i^+))$$

$$\begin{aligned}\ln\hat{R} &= \frac{N_u}{\left(\frac{\beta_0}{q_0}\right)^\gamma \cdot \sum_{\delta_i=1}\ln(1-F_0(T_i^+)) - N_c\lambda^\gamma} \\ \Rightarrow \ln\hat{R} &= \frac{-N_u}{\lambda^\gamma \cdot \sum\left(\frac{T_i^+}{L}\right)^\gamma} \end{aligned} \tag{6.3}$$

Theorem 8
The MLE of CUS sampling, given by

$$\ln\hat{R} = \frac{-N_u}{\lambda^\gamma \cdot \sum\left(\frac{T_i^+}{L}\right)^\gamma},$$

has the same form as the MLE of the sampling distribution $T_i \sim \mathcal{W}(\beta;\gamma)$ for known γ but with T_i replaced by T_i^+.

Proof *Derive the MLE for $\hat{\beta}$ from $\hat{R}$ using the Weibull CDF:*

$$\begin{aligned}\ln\hat{R} &= -\left(\frac{q_0}{\hat{\beta}}\right)^\gamma \\ \Leftrightarrow \frac{N_u}{\lambda^\gamma \cdot \sum\left(\frac{T_i^+}{L}\right)^\gamma} &= \left(\frac{q_0}{\hat{\beta}}\right)^\gamma \\ \Leftrightarrow \hat{\beta}^\gamma &= q_0^\gamma \frac{\lambda^\gamma \cdot \sum\left(\frac{T_i^+}{L}\right)^\gamma}{N_u} \\ &= \frac{\sum\left(T_i^+\right)^\gamma}{N_u}\end{aligned}$$

Looking at the MLE (B.1) of appendix B gives the claim.

qed

6.3 Confidence Intervals from CUS Sampling

CUS sampling decides about a hypothesis $H_0: F(q_0) \le p_0$ with significance α. For a given sample $\underline{T}^+$, $\underline{\delta}$, define the CUS-CI for R as the set of all H_0 which could be rejected:

$$\{R \mid B_L \ge m_{\text{crit}},\ H_0: F(q_0) \le 1-R\}$$

Here, $F(q_0) = 1-R$ determines $\beta_0 = q_0 \cdot (-\ln R)^{-\frac{1}{\gamma}}$. It follows for the CI:

$$\begin{aligned} m_{\text{crit}}(R) &\le \sum_{i=1}^N F_0(T_i^+) = N - \sum_{i=1}^N e^{-\left(\frac{T_i^+}{\beta_0}\right)^\gamma} \\ &= N - \sum_{i=1}^N e^{-\left(\frac{T_i^+}{q_0\cdot(-\ln R)^{-\frac{1}{\gamma}}}\right)^\gamma} \\ &= N - \sum_{i=1}^N R^{\left(\frac{T_i^+}{q_0}\right)^\gamma}\end{aligned}$$

Numerical solving of

$$\sum_{i=1}^{N} R^{\left(\frac{T_i^+}{q_0}\right)^\gamma} + m_{\text{crit}}(R) \leq N$$

gives the desired CI.

6.4 Summary

Calculating likelihood CUS sampling for Weibull distributions does lead to a sum of censored beta distributions. A MLE can be obtained via

$$\ln \hat{R} = \frac{-N_u}{\lambda^\gamma \cdot \sum \left(\frac{T_i^+}{L}\right)^\gamma}$$

being equivalent to a MLE for fixed γ using the sampling distributions likelihood.

By construction, the CUS CI gives correct significance when used for testing. However their application requires a fixed shape parameter. In this sense, **CUS sampling is a hybrid of reliability estimation and reliability demonstration**. The test statistic is a direct generalisation of success runs. From the test statistics likelihood, a point estimator can be defined, equivalent to MLE for the sampling distribution. Different to other quantile estimates the method is free of coverage errors or discontinuities.

Chapter 7

Bayesian Reliability Demonstration

To account for previous knowledge in binomial sampling[1] Bayesian methods for success runs as described in [26], [7] and [8] can be used. After a short introduction to the current theory the methods are applied to CUS sampling (For an introduction on Bayesian statistics see appendix D).

7.1 Bayesian Methods for Binomial Sampling

7.1.1 Uniform Priors

Consider again the setup of success runs of chapter 3. If all tested units survive test duration L the samples likelihood, conditional to H_0, is:

$$L(T_1 = \cdots = T_N = L \mid H_0) = (1 - F_{\theta_0}(L))^N$$

where θ_0 is uniquely determined by H_0. Homologation is granted if:

$$(1 - F_{\theta_0}(L))^N \leq \alpha$$

Using the term **reliability** $R_L = 1 - F_{\theta_0}(L) = 1 - p_L$ practitioners sometimes interpret R_L^N as a p-value. According to statistical test theory H_1 is accepted if the p-value falls below α.

Remark 13:
German fatigue literature[2] often uses the term "Aussagesicherheit", translating to "statement certainty" when speaking about the p-value and writes $P_A = R_L^N$. This notion suggests, that P_A is the probability of drawing a correct conclusion from the test, while it is actually the significance. Therefore this term has to be avoided!

For one-parameter distribution models there is a 1:1-correspondence between R_L and the unknown scale parameter. The Bayesian approach of [4][3] now uses

[1] See also paradigm **P2**
[2] e.g. [3]
[3] see also [2]

a uniform prior[4] for the reliability R_L with prior density:

$$\pi(R_L) = \begin{cases} 1, & 0 \leq R_L \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Using the Bayesian theorem (D.1) gives the posterior density:

$$\pi(R_L \mid T_1 = \dots = T_N = L) = \frac{R_L^N}{\int_0^1 R_L^N \, dR_L} = (N+1)R_L^N \tag{7.1}$$

Constructing a left sided $(1-\alpha)$-CI for R_L gives:

$$\begin{aligned} 1-\alpha &= \int_{R_{\text{CI}}}^{1} \pi(R_L \mid T_1 = \dots = T_N = L) \, dR_L \\ \Leftrightarrow \alpha &= R_{\text{CI}}^{N+1} \end{aligned} \tag{7.2}$$

Comparing equation (7.2) with the common Weibull significance equation $R_L^N = \alpha$ gives rise to the idea, that using the uniform prior reduces the necessary sample size by one unit.

Remark 14 (Uniform prior paradox):
Some fatigue literature[5] calls this approach "success run with presupposition", reflecting that R_L is a probability, i.e. $R_L \in [0,1]$. The possibility of lowering the necessary sample size by one sounds paradox: Even though π seems to model no new information, the sample size is reduced.

This prior gives $H_0 : R_0 \leq 1 - p_0$ a large prior probability of $1 - p_0$, making a success run strongly indicate H_1. Therefore the "presupposition" of the approach is not that R_0 is known to be in $[0,1]$, but in fact the strong supposition that H_0 is true. It is now comprehensible, that less efforts have to be spend to topple this supposition.

However, R_L is also a probability, so why not applying the same uniform prior argument to R_L? In this case H_0 would get a small prior probability $1-p_L$. Therefore this paradox originates from the missing invariance of Bayesian methods under reparametrisation[6].

7.1.2 Beta Priors for the Case $L = q_0$

In [4] a compound prior, consisting of a uniform and a beta[7] prior, is introduced:

$$\pi(R_0) = \frac{\rho}{B(A_0, B_0)} R_0^{A_0-1}(1-R_L)^{B_0-1} + (1-\rho) \tag{7.3}$$

The factor ρ determines the contribution of the beta distribution to the prior and is called **knowledge factor**, again emphasing a uniform prior to contain no information (which is not true due to remark 14).

[4] In fact the uniform distribution is a special case of the beta distribution, see remark 20
[5] e.g. [3] or [2]
[6] See [25] for more information about uniform "noninformative" priors
[7] see section A.4

Remark 15:
The prior introduced by (7.3) contains the special cases of a uniform prior ($\rho = 0$) and a pure beta prior ($\gamma = 1$), used to summarise results from previous similar experiments. Hence ρ is a measure for the transferability of previous knowledge. **Since uniform distributions belong to the beta family, it is sufficient to study this family's behaviour when used for priors:**

$$\pi(R_0) = \frac{1}{B(A_0, B_0)} R_0^{A_0-1}(1-R_0)^{B_0-1} \tag{7.4}$$

Observing S_{q_0} passed components w.r.t time L gives the likelihood

$$L(R_0 \mid S_{q_0}) = \binom{N}{S_{q_0}} R_0^{S_{q_0}}(1-R_0)^{N-S_{q_0}}$$

Using Bayesian formula gives the posterior density:

$$\begin{aligned}
\pi(R_0 \mid S_{q_0}) &= \frac{\binom{N}{S_{q_0}} R_0^{S_{q_0}}(1-R_0)^{N-S_{q_0}} \frac{1}{B(A_0,B_0)} R_0^{A_0-1}(1-R_0)^{B_0-1}}{\int\limits_0^1 \binom{N}{S_{q_0}} R^{S_{q_0}}(1-R)^{N-S_{q_0}} \frac{1}{B(A_0,B_0)} R^{A_0-1}(1-R)^{B_0-1}\, dR} \\
&= \frac{R_0^{(A_0+S_{q_0})-1}(1-R_0)^{(B_0+N-S_{q_0})-1}}{\int_0^1 R^{(A_0+S_{q_0})-1}(1-R)^{(B_0+N-S_L)-1}\, dR} \qquad (7.5)
\end{aligned}$$

Hence, $\pi(\cdot \mid S_{q_0})$ is the PDF of a Beta$(A_0 + S_{q_0}; B_0 + N - S_{q_0})$ distribution. I.e. the **beta distribution is a conjugated family for success runs with $L = q_0$ and parameter update formula**

$$A = A_0 + S_{q_0},\ B = B_0 + (N - S_{q_0}) \tag{7.6}$$

In this context we call:

- A_0: Pseudo number of survivors
- $A_0 + S_{q_0}$: Combined number of survivors
- $A_0 + B_0$: Pseudo sample size
- $A_0 + B_0 + N$: Combined sample size

A Bayesian version of the significance equation (5.7) can be formulated as:

$$\int\limits_{R_0}^{1} \pi(R \mid S_{q_0}) dR = 1 - \alpha$$

Using formula (A.3) gives:

$$\sum_{j=A_0+S_{q_0}}^{A_0+B_0+N-1} \frac{(A_0+B_0+N-1)!}{j!(A_0+B_0+N-1-j)!} R_0^j (1-R_0)^{A_0+B_0+N-1-j} = \alpha$$

This equation can in general not be solved analytically, hence numerics are used whenever necessary.

Remark 16:
Some hints for choosing adequate priors can be found in [26] section 6.5. There, the prior construction is based on its desired 5%-, 95%-quantile and its mean.

7.1.3 Priors for the Case $L > q_0$

If two tests differ in their values for L, it might be still desirable to use the same prior. Therefore it is not advisable to simply use the methods from the previous section to model knowledge about R_L, but to formulate priors for R_0:

$$\pi(R_0) = \frac{1}{B(A_0, B_0)} R_0^{A_0-1}(1 - R_0)^{B_0-1}$$

For the remaining chapter a Weibull distribution with shape γ is assumed, to get an easy relation between the reliabilities[8] R_0 and R_L:

$$R_L = R(q_0)^{\lambda^\gamma},\ \lambda = \frac{L}{q_0}$$

The likelihood does now assyme the form:

$$L(R \mid S_L) = \binom{N}{S_L} \left(R^{\lambda^\gamma}\right)^{S_L} (1 - R^{\lambda^\gamma})^{N-S_L}$$

From the last factor it can immediately be seen, that the current prior π and the likelihood L will in general not lead to a conjugated family. **Only in case of $N = S_L$ the factor $(1 - R^{\lambda^\gamma})$ vanishes and the posterior is a beta PDF:**

$$\boxed{\pi(R_0 \mid S_L = N) = \frac{1}{\beta(A_0 + N\lambda^\gamma, B_0)} R_0^{A_0+N\lambda^\gamma}(1-R_0)^{B_0}}$$

7.1.4 Summary

When using binomial sampling, the class of beta distributions is a conjugated family for $L = q_0$. The parameter update is given by the simple formula

$$A = A_0 + S_{q_0},\ B = B_0 + N - S_{q_0},$$

where S_{q_0} is the number of component surviving time $L = q_0$.

For the case $L > q_0$ the update formula for Weibull distributions is

$$A = A_0 + N\lambda^\gamma,\ B = B_0,$$

if $S_L = N$. Otherwise, for $S_L < N$ (likely for large L) no update formula is available.

[8] See section A.2.3

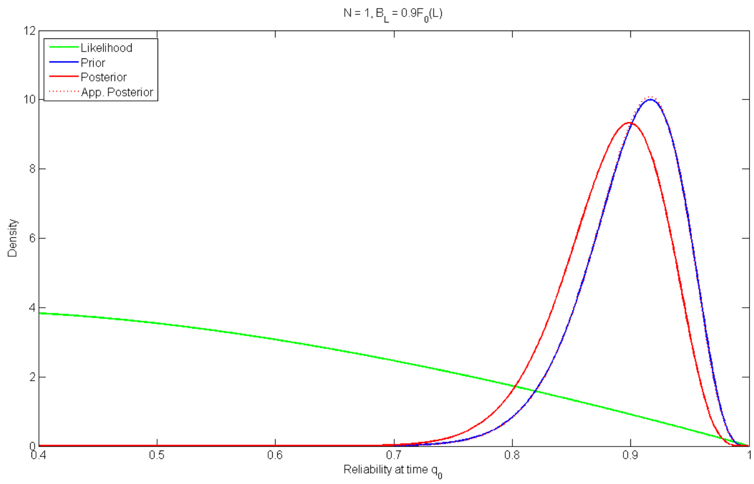

Figure 7.1: Prior approximation for $N = 1$ and a result $B_L = 0.9 \cdot F_0(L)$.

7.2 Bayesian Methods for CUS Sampling

In plain language, PPC counts are success runs with non-integer values for S_L, suggesting that beta priors might at least be approximately conjugated to CUS sampling. We will now adapt beta priors to CUS sampling to get an approximative beta posterior PDF for the case $L > q_0$ and $S_L < N$.

7.2.1 Transferring the Conjugated Family

Let π be a beta prior with parameters A_0, B_0. Our first guess of generalising the update formula (7.6) for $L = q_0$ to CUS sampling would be

$$A = A_0 + \frac{1}{p_L} B_L,\ B_0 = B_0 + (N - \frac{1}{p_L} B_L)$$

The following example shows the falseness of this guess.

Example 14:
Let $N = 1$ and $L = q_0 = 10,000$ for a Weibull distribution with $\gamma = 2$ and $p_0 = 0.01$. This test will not fulfill the significance equation, but we are only interested in the prior approximation. If the prior was chosen to be Beta$(45; 5)$ and the result was $B_L = 0.9 \cdot F_0(L)$, then we try to approximate the posterior by a Beta$(45 + 0.9; 5 + (1 - 0.9))$. Figure 7.1 shows that approximation is not even satisfying, since the approximated posterior is closer to the prior than to the true posterior.

7.2.2 Derivation of an Approximation Formula

To find the correct update formula we start with the true posterior. $\bar{\pi}$ can be calculated in N steps using lemma 9. This means we can stick to $N = 1$ w.l.o.g.:

$$\bar{\pi}(R) \propto R^{A_0-1}(1-R)^{B_0-1} \cdot \left(g_1\left(u\right) \cdot \mathbb{1}_{[0,F_0(L)]}(u) + (1-F(L))\delta_{F_0(L)}(u)\right)$$

g_1 is the PDF of a Beta$(1; B(R))$ with $B(R) = \left(\frac{\beta_0}{\beta(R)}\right)^\gamma = \frac{-\log R}{-\log R_0}$. Further $1 - F(L) = R^{\lambda^\gamma}$ gives:

$$\begin{aligned} \bar{\pi}(R) \quad \propto \quad & R^{A_0-1}(1-R)^{B_0-1}B(R)\frac{R^{\frac{-\ln(1-u)}{-\ln R_0}}}{1-u}\mathbb{1}_{[0,F_0(L)]}(u) \\ & + R^{A_0+\lambda^\gamma-1}(1-R)^{B_0-1}\delta_{F_0(L)}(u) \end{aligned} \quad (7.7)$$

Using a Taylor expansion for the logarithm gives $\log s \approx s - 1$, for $s \approx 1$. If $\pi(R) \approx 0$ for $R \ll 1$, then also $\ln R \approx R - 1$. It follows:

$$\begin{aligned} (7.7) \quad \approx \quad & R^{A_0+\frac{-\ln(1-u)}{-\ln R_0}-1}(1-R)^{B_0+1-1}\frac{1}{(-\ln R_0)\cdot(1-u)}\mathbb{1}_{[0,F_0(L)]}(u) \\ & + R^{A_0+\lambda^\gamma-1}(1-R)^{B_0-1}\delta_{F_0(L)}(u) \end{aligned}$$

We now have two different possibilities for parameter update:

$$T^+ < L \quad \Rightarrow \quad A = A_0 + \frac{-\ln(1-F_0(T^+))}{-\ln R_0},\ B = B_0 + 1 \quad (7.8)$$

$$T^+ = L \quad \Rightarrow \quad A = A_0 + \lambda^\gamma,\ B = B_0 \quad (7.9)$$

This means, that partially-passed components will always add one pseudo failure and a partial pseudo survivor.

Lemma 6 *Let $T_1^+, \ldots, T_N^+$ be censored at time $L = \lambda q_0$, $\lambda \geq 1$ with censoring indicator δ_i and $F_0 \sim \mathcal{W}(\beta_0; \gamma_0)$ and $\pi \sim$ Beta$(A_0; B_0)$ be a prior for $R = 1 - F(L)$, where F is the sampling distribution of T_i, then the posterior $\bar{\pi}$ can be approximated by a beta distribution* Beta$(A; B)$ *with:*

$$\begin{aligned} A &= A_0 + \lambda^\gamma \sum \left(\frac{T_i^+}{L}\right)^\gamma \\ B &= B_0 + N - \sum \delta_i \end{aligned}$$

Proof *Apply lemma 9 consecutively for $T_1^+, \ldots, T_N^+$. Each time $\delta_i = 0$ occurs equation (7.8) applies, adding summand $\frac{-\ln(1-F_0(T^+))}{-\ln R_0}$ to A and summand 1 to B. Each time $\delta_i = 0$ occurs equation (7.9) applies, adding summand λ^γ to A.*

Finally Weibulls quantile property can be used to get

$$\frac{-\ln\left(1-F_0\left(T_i^+\right)\right)}{-\ln R_0} = \left(\frac{T_i^+}{q_0}\right)^\gamma = \lambda^\gamma\left(\frac{T_i^+}{L}\right)^\gamma$$

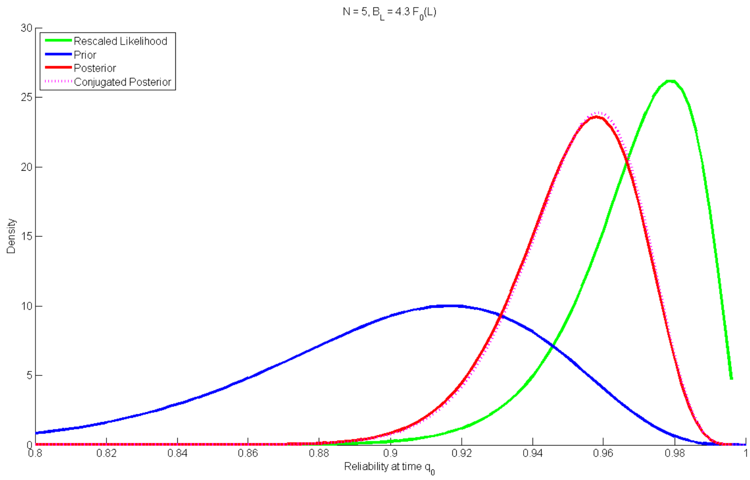

Figure 7.2: Prior approximation for $N = 5$ and a result $B_L = 4.3 \cdot F_0(L)$.

and

$$\sum_{\delta_i=0} \left(\frac{T_i^+}{L}\right)^\gamma + (\sum \delta_i) = \sum \left(\frac{T_i^+}{L}\right)^\gamma$$

qed

Example 15:
Let $L = 5 \cdot q_0$ for a Weibull model with $\gamma = 2$. Assume that $N = 5$ and $F_0(T_1^+) = \cdots = F_0(T_3^+) = F_0(L)$, $F_0(T_4^+) = 0.8 \cdot F_0(L)$, $F_0(T_4^+) = 0.5 \cdot F_0(L)$. Hence $B_L = 4.3F_0(L)$ and $S_L = 3$. If $R_0 = 0.9$ has to be verified, and a prior $\pi \sim \text{Beta}(45; 5)$ is used, then the posterior can be calculated being $\text{Beta}(139.81; 7)$. Figure 7.2 shows, that the approximation is very good.

Remark 17:
Different to the test statistic B_L of CUS sampling, the posterior approximation (as well the MLE of CUS sampling) does depend on the detailed observations T_i^+. The next lemma shows, that there can not be a different approximation strategy using only B_L.

Lemma 7 *The statistic B_L (depending only on the hypothesis, not on the sampling distribution) is not sufficient.*

Proof *Show that the sufficiency principle*[9] $L(R \mid \underline{T}^+, \underline{\delta}) \propto L(R \mid B_L)$ *is violated. Consider a result* $F_0(T_1^+) = s \cdot p_L$, $F_0(T_2^+) = (1-s)p_L$, $s \in [0,1]$, *always leading to* $B_L = 1 \cdot p_L$. *Due to Theorem 6:*

$$L(R \mid sp_L,\ (1-s)p_L) = B^2(R)\left((1-sp_L)(1-(1-s)p_L)\right)^{B(R)-1}$$

It can immediately be seen, that L does depend on s:

$$(1-sp_L)(1-(1-s)p_L) = 1-sp_L-(1-s)p_L+s(1-s)p_L = 1-p_L+s(1-s)p_L \not\equiv \text{const}$$

[9] See [27] section 2.9.

qed

7.3 Bayesian Design of Experiments

7.3.1 Priors Using a Knowledge Factor

Success runs and PPC counts as introduced in chapters 5 and 3 require the same minimal sample size:

$$N_{\min} = \left\lceil \frac{\ln \alpha}{\ln(1-p_L)} \right\rceil$$

In this section we will study the influence of the knowledge factor (see equation (7.3)) on the sample size and test duration, i.e. consider priors of the form

$$\pi(R_0) = \frac{\rho}{B(A_0,B_0)} R_0^{A_0-1}(1-R_L)^{B_0-1} + (1-\rho) = \rho\pi_\beta + (1-\rho)\pi_{\text{uniform}}$$

The case $\rho = 0$ leads to a uniform prior as described in section 7.1.1, leading to sample size $N_{\min} - 1$. Once a priors beta component offers

$$\int_{R_0}^{1} \bar{\pi}_\beta(R)dR \le \alpha \tag{7.10}$$

no additional sample would be required, i.e. $N = 0$ for $\rho = 1$.

7.3.2 Derivation of a Design Formula

Let us consider a prior π with beta component fulfilling equation (7.10) and $0 < \rho < 1$. Design of experiments (for success runs) consists of solving

$$\mathbb{P}_{H_0}(S_L = N \mid \bar{\pi}) \le \alpha \tag{7.11}$$

where the posterior $\bar{\pi}$ is calculated as:

$$\begin{aligned} \bar{\pi}(R) &\propto \pi(R) f_{\underline{T}}(R) = (\rho\pi_\beta(R) + (1-\rho)\pi_{\text{uniform}}(R))\, f_{\underline{T}}(R) \\ &= \rho\pi_\beta(R) f_{\underline{T}}(R) + (1-\rho)\pi_{\text{uniform}}(R) f_{\underline{T}}(R) \end{aligned}$$

The summands $\pi_\beta(R) f_{\underline{T}}(R)$ and $\pi_{\text{uniform}}(R) f_{\underline{T}}(R)$ are proportional to posteriors given by lemma 6. Unfortunately, the proportionality coefficients will differ, such that $\bar{\pi}(R)$ is in fact a linear combination of the single posteriors, but not with weights ρ, $1-\rho$. The true posterior has to be calculated numerically.

Lemma 6 shows, that $\bar{\pi}$ does only depend on the success runs result N, λ via $N\lambda^\gamma$. This means that solving equation (7.11) gives a result:

$$N \cdot \lambda^\gamma \ge f_\rho$$

Notice, that a classical success run (as described for Weibull distributions in section 3.4.2) would give the design formula

$$N \cdot \lambda^\gamma \ge \frac{\ln \alpha}{\ln R_0}$$

N	1	2	3	4	5
λ	10.8	7.3	5.9	5	4.4

Table 7.1: Bayesian DoE for verifying a reliability of 99% with 1% significance.

The Bayesian design parameter f_ρ can be calculated numerically (depending on R_0, α, ρ and the parameters of the priors beta component), as the following example indicates.

Example 16:
Let π_β be the density of a Beta(1200; 5) and $\rho = 0.6$, then we have:

$$\Pi_\beta(0.99) = 0.0072, \Pi(0.99) = 0.2037$$

This mean, that the reliability described by π_β was satisfying (if $R_0 = 0.99$ and $\alpha = 0.01$ was claimed), but after a 60% knowledge transfer homologation is doubtfull. From example 11 it is known, that $N = 5$, $\lambda = 12.5$ is feasible for a Weibull distribution with $\beta = 1.8$.
For different N the value λ was chosen such that the posteriors (calculated numerically) fulfills

$$\int_{R_0}^{1} \bar{\pi}(R)dR \leq 0.99$$

The results can be seen in table 7.1, showing that $N \cdot \lambda^\gamma \geq 72.3429$. Success runs claim $N \cdot \lambda^\gamma \geq 458.21$, which means a factor of 6 in the efforts.

7.4 Summary

Bayesian methods for binomial sampling use beta distributions as a conjugated family. If $L = q_0$, then any (integer) result S_L will lead to $\bar{\pi}$ being a beta distribution. If $L > q_0$ then only $S_L = N$ will lead to beta posteriors.

When switching to CUS sampling, the true posterior is not exactly a beta PDF, but the approximation is very good. The parameter update formula does also work for $L > q_0$ for any result. In this sense, CUS sampling generalises the property of binomial samplings conjugated beta family to any result and test duration.

Therefore, the success runs second drawback of section 1.1 did not transfer to PPC counts.

Chapter 8

Power of RET and RDT Methods

It is known from lemma 3 and section 5.4, that randomisation and CUS sampling do not show the conservative behaviour of binomial sampling in the area of $p \approx p_0$. Further section 2.5.3 has shown that the corrected Delta also improves the power of the ordinary delta method. It is left to compare the power of randomisation, CUS sampling and the corrected delta method, as well as the robustness of all methods against misspecified distribution models.

8.1 Power Calculation using Monte Carlo

Calculating the power of CUS sampling and the corrected delta method is done by Monte Carlo simulation. For each point $p = F(q_0)$ the corresponding parameter θ_p can be calculated (one parameter distribution model), allowing to simulate M samples of size N from the distribution F_{θ_p}.

Let X denote the number of rejected hypotheses out of M simulations, then $\frac{X}{M}$ is an estimate for the power functions value belonging to p:

$$\hat{g}(p) = \frac{X}{M}$$

Notice that X has a binomial distribution with probability of success $g(p)$.

Near p_0, the values of the power function are expected to be around the significance, which is here chosen to be $\alpha = 0.01$. If we claim to have a variation coefficient of about 5% for $\hat{g}_{L,\mathrm{CUS}}(p_0)$, then for the simulation size M it holds:

$$\begin{aligned} \frac{\sqrt{\mathbb{V}\mathrm{ar}(X/M)}}{\mathbb{E}(X/M)} \approx \sqrt{\frac{1-p_0}{Mp_0}} &\leq 0.05 \\ \Leftrightarrow \frac{1-p_0}{0.05^2 \cdot p_0} &\leq M \end{aligned}$$

For $p_0 = 0.01$ a simulation size of $M \geq 39,600 \approx 40,000$ is required.

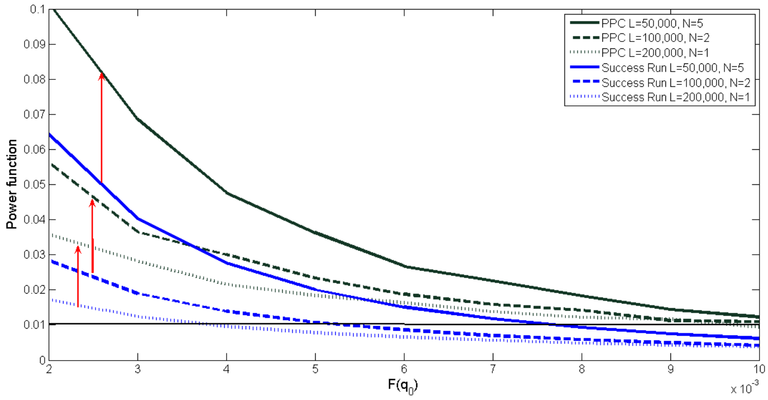

Figure 8.1: PPC count power function for different L with $p_0 = 0.01$, $q_0 = 10,000$, $\alpha = 0.01$, lognormal with $\sigma^2 = 0.36$.

8.2 The Power of CUS Sampling

Using Theorem 6 allows to calculate the power $g_{\text{CUS}}(p)$. Hereto, the PDF g_N^+ of B_L is obtained as the N-fold convolution of:

$$g_1^+(u) = g_1(u) \cdot \mathbb{1}_{[0,F_0(L)]}(u) + (1 - F(L))\delta_{F_0(L)}(u)$$

where the parameters of the beta PDF g_1 depend on the reliability R.

The power can now be written as:

$$g_{\text{CUS}}(p) = 1 - G_N^+ (m_{\text{crit}}(R_0)), \; G_N^+ \text{ depending on } p$$

Unfortunately, there is no easy formula for convolutions of g_1^+, since sums of beta variables follow no easy model[1].

Example 17:
Consider a lognormal distribution with $\sigma^2 = 0.36$ and design three tests for $p_0 = 0.01$, $q_0 = 10,000$, $\alpha = 0.01$ using the pairs (N, L): $(5, 50000)$, $(2, 100000)$ and $(1, 200000)$. While all three offer a consumers risk below α, they differ in their producers risk, as was shown in figure 3.1 (and in figure 8.1).

The critical values for the PPC count are 4.9061, 1.9572, 0.99398 and the power[2] (in the area of H_1) of the three designs are shown in figure 8.1. The power does still depend on the special experimental design, but does not longer behave conservative, i.e. now: $g(p_0) = \alpha$, and is raised as indicated by the red arrows.

The first drawback of success runs of section 1.1 is now reduced: **The power of CUS sampling does still depend on the special choice of the design (N, L) but offers the correct significance**.

[1] In [12] the problem is solved using the CLT, which will not work for our small sample sizes

[2] Monte Carlo simulation with 40,000 samples per point

8.3 Power of Randomisation of CUS Sampling

8.3.1 Correct Distribution Models

Theoretical Arguments

CUS sampling was constructed for improving the binomial sampling power by removing discreteness from the test statistic.

Theorem 9
Let $F(q_0) = p$ be the true failure quota at time q_0 and $H_0 : p \geq p_0$, and let $g_{L,\text{CUS}}$, $g_{L,\text{BIN}}$ denote the power functions of the CUS and binomial models resp., then it holds:

$$\forall p \in [0,1] : \; g_{L,\text{CUS}}(p) \geq g_{L,\text{BIN}}(p)$$

Further, typically $g_{L,\text{CUS}}(p_0) = \alpha$, $g_{L,\text{CUS}} < \alpha$, where α is the given significance.

Proof *First by definition: $\frac{1}{p_L} B_L \geq S_L$, not depending on p. Secondly $S_{\text{crit}} \in \mathbb{N}$ such that:*

$$\alpha = \int_{S_{\text{crit}}}^{NF_{\vartheta_0}(L)} h_N(u)du$$

can in general not be solved for integer S_{crit}, hence $S_{\text{crit}} > m_{\text{crit}}$ to ensure $g_{L,\text{BIN}} \leq \alpha$. This means that $S_L \geq S_{\text{crit}}$ will always imply $\frac{1}{p_L} B_L \geq m_{\text{crit}}$. Inversely samples of the form

$$T_1 = \ldots = T_{N-1} = L, \; T_N = qL$$

show for sufficiently high $q < 1$ that the the reverse is not true for success runs.

qed

From lemma 3 we also know, that

$$\forall p \in [0,1] : \; g_{L,\text{Rand}}(p) \geq g_{L,\text{BIN}}(p)$$

Since randomisation and CUS sampling both have exactly the correct significance, it is left to study which one has the better power on the rest of the interval $[0,1]$.

Theorem 10
Let $N = 1$ and L such that the significance equation (3.6) holds, then for a Weibull model:

$$\forall p \in [0,1] : \; g_{L,\text{Rand}}(p) \leq g_{L,\text{CUS}}(p)$$

where equality does hold if and only if equality holds in (3.6).

Proof *Let $p_L(p)$ denote the failure probability at time L, if p is the true failure probability at time q_0. From equations (4.2) and (4.3) under $N = 1$ we get the power of randomisation as:*

$$\begin{aligned} g_{L,\text{Rand}}(p) &= \left(1 - \frac{\alpha - (1 - p_L(p_0))}{p_L(p_0)}\right)(1 - p_L(p)) + \frac{\alpha - (1 - p_L(p_0))}{p_L(p_0)} \\ &= 1 - \frac{1-\alpha}{p_L(p_0)} \cdot p_L(p) \end{aligned}$$

If the significance equation (3.6) is fulfilled, $N = 1$ gives $1 - p_L(p_0) \leq \alpha$, hence:

$$\begin{aligned} g_{L,\mathrm{Rand}}(p) &= 1 - \frac{1-\alpha}{p_L(p_0)} \cdot p_L(p) \\ &\leq 1 - 1 \cdot p_L(p) \\ &= R_L(p) = R_0(p)^{\lambda^\gamma} = (1-p)^{\lambda^\gamma} = \alpha^{-\ln(1-p)\cdot\left(\frac{\beta_0}{q_0}\right)^\gamma} \\ &= \alpha^{B(1-p)} \end{aligned}$$

Under H_0 B_L has for $N = 1$ the PDF $f(s) = \mathbb{1}_{[p_L(p_0),1]}(s) + (1 - p_L(p_0)) \cdot \delta_{p_L(p_0)}(s)$ (see section C.2), giving the critical value as

$$F(m_{\mathrm{crit}}) \stackrel{!}{=} 1 - \alpha$$

From lemma 5 it follows that B_L has CDF:

$$G_1^+(s) = 1 - (1-s)^{B(R)},\ B(R) = -\left(\frac{\beta_0}{q_0}\right)^\gamma \cdot \ln R = \frac{\ln R}{\ln R_0}$$

It follows:

$$g_{L,\mathrm{CUS}}(p) = 1 - G_1^+(m_{\mathrm{crit}}) = \alpha^{B(1-p)} \geq g_{L,\mathrm{Rand}}(p)$$

qed

8.3.2 Exemplary Power Calculation

Example 18:
Consider again a lognormal distribution with $\sigma = 0.6$ and $p_0 = 0.01$, $q_0 = 10{,}000$, $\alpha = 0.01$, $L = 5q_0$. From example 10 it is known that $p_L = 0.9349$ and $\phi_B = 0.0735$. Since there is no analytic formula for the power function of the PPC count available, Monte-Carlo-Simulation is used. Fig. 8.2 shows, that the PPC count does not only have the correct significance, but has also approximately the same power as randomisation. The critical value of $p_L^{-1} B_L$ was calculated to be $m_{\mathrm{crit}} = 4.9357$. If e.g. four units pass time L, the fifth one has to achieve about 0.94 points, which means a lifetime of $F_{\beta_0}^{-1}(p_L \cdot 0.94) \approx 47061$.

In this case, the difference between the power functions of CUS and binomial models is about 2% in the domain of H_1. This means: If 100 decisions about homologations of feasible components are made, then the binomial models will have 2 more cost-intensive wrong-negative decisions, than CUS models will have.

While the power functions of the CUS model and randomisation are nearly the same, the CUS model is fully repeatable since its test statistic and critical value are deterministic. The additionally clear interpretability should give the PPC count a much higher acceptance in practice.

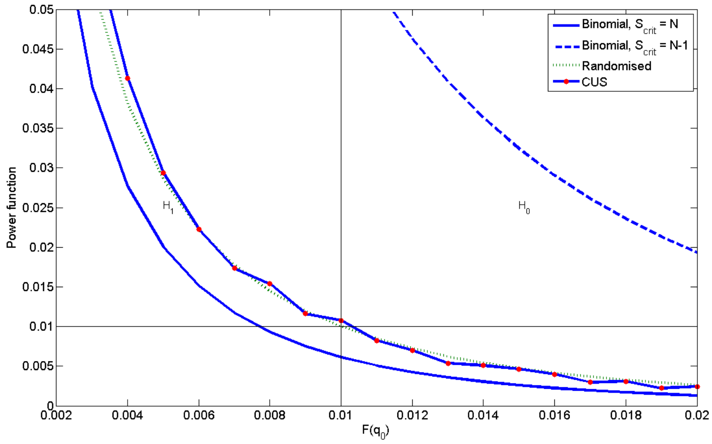

Figure 8.2: Power functions for $p_0 = 0.01$, $q_0 = 10,000$, $\alpha = 0.01$, $L = 5q_0$, for a lognormal distribution with $\sigma^2 = 0.36$. Sample sizes $N = 5$ due to table 3.1.

8.4 Misspecified Distribution Models

To get one-parameter distribution models, the shape parameter of Weibull and lognormal distributions was fixed. What if this fixing was not suitable?

Example 19 (Example 18 continued):
The power functions will be calculated again, but this time the fixed value $\sigma^2 = 0.36$ will not be the true one. The critical values and test statistics are not effected by the true distribution, only the power functions:

- If the true σ^2 is smaller than the fixed one, like in figure 8.3, all models show an increased conservative behavior.
- If the true σ^2 is larger, like in figure 8.4, all tests fail to keep the significance.

In both cases randomisation and CUS sampling show approximately the same power.

The robustness in the upper example can be explained by looking at the dependence of test statistics and critical values on the fixed parameters. First S_L does not depend on the fixation, different to the critical value S_{crit}:

- If the fixed parameter underestimates the true variance, too few units are tested, giving an increased actual significance.
- If the fixed parameter overestimates the true variance, too many units are tested, giving a more conservative test.

Since ϕ_B does not depend on the fixation as well, randomisation looses power in the same degree as binomial sampling. The PPC count and its critical value m_{crit} depend on the distribution model, but in combination the result of the test shows the same dependence on the model as randomisation.

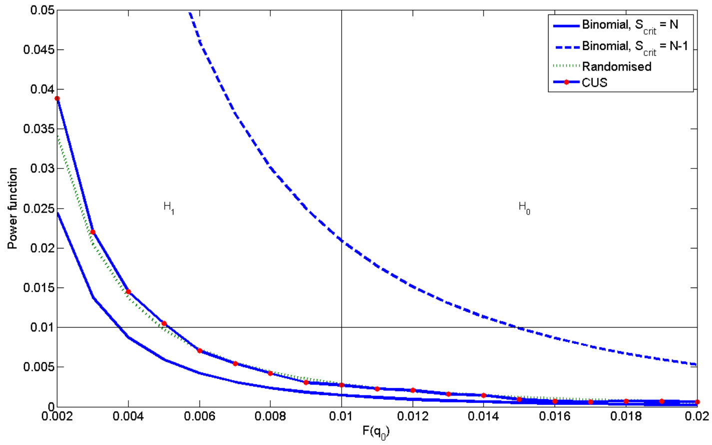

Figure 8.3: Power functions for $p_0 = 0.01$, $q_0 = 10,000$, $\alpha = 0.01$, $L = 5q_0$, for a lognormal distribution with $\sigma^2 = 0.30$. **Assumed**: $\sigma^2 = 0.36$. Sample sizes due to table 3.1. 20,000 samples per MC simulation.

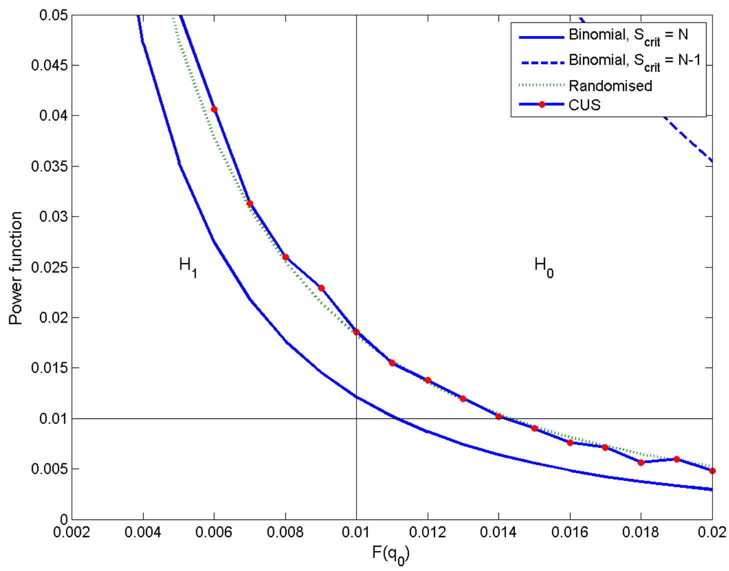

Figure 8.4: Power functions for $p_0 = 0.01$, $q_0 = 10,000$, $\alpha = 0.01$, $L = 5q_0$, for a lognormal distribution with $\sigma^2 = 0.40$. **Assumed**: $\sigma^2 = 0.36$. Sample sizes due to table 3.1. 20,000 samples per MC simulation.

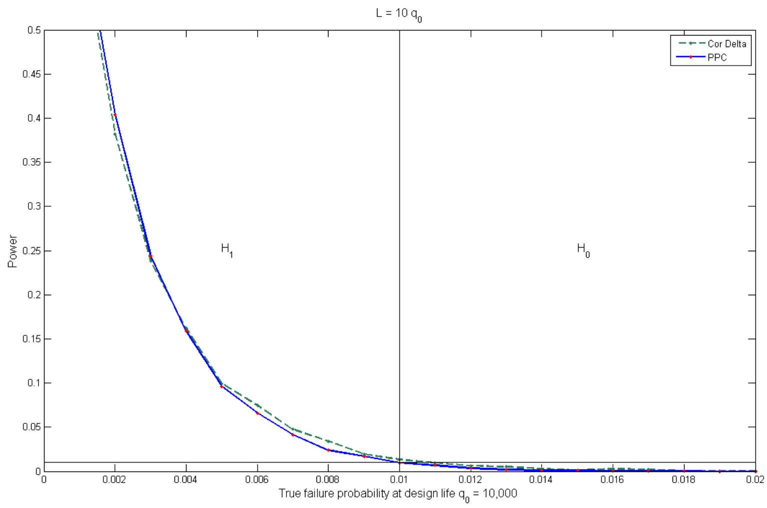

Figure 8.5: Power functions of corrected delta and PPC count for $L = 10q_0$.

Remark 18:
Example 19 shows, that σ^2 should be fixed conservatively (rather overestimated than underestimated), in order to keep the significance. In general, it is better to fix the shape parameter such that the resulting variance of the model is greater or equal to the true variance. For Weibull models, the shape γ should be bounded from above, since the variance increases as γ decreases.

8.5 Power of CUS Sampling and Corrected Delta

To close the section, we want to show that CUS sampling has convenient benefits towards the corrected delta method. If H_1 is true, most homologations from the delta method happen when the sample is completely censored (i.e. one switches to success runs). As soon as failures are present, quantile estimates drop discontinuously to smaller values, as in figure 2.3. Even though this was fixed by lowering the critical value, the whole test is not very convincing, but appears to be a technical artifice.

CUS sampling is free of such discontinuous test statistics and is a more elegant technique. Instead of switching between quantile estimates and success runs, one stays in one concept having an easy interpretation and no coverage error. We believe in a good acceptance in practice.

Example 20 (Example 5 continued):
Figures 8.5 and 8.6 show that both methods are practically equivalent in terms of power.

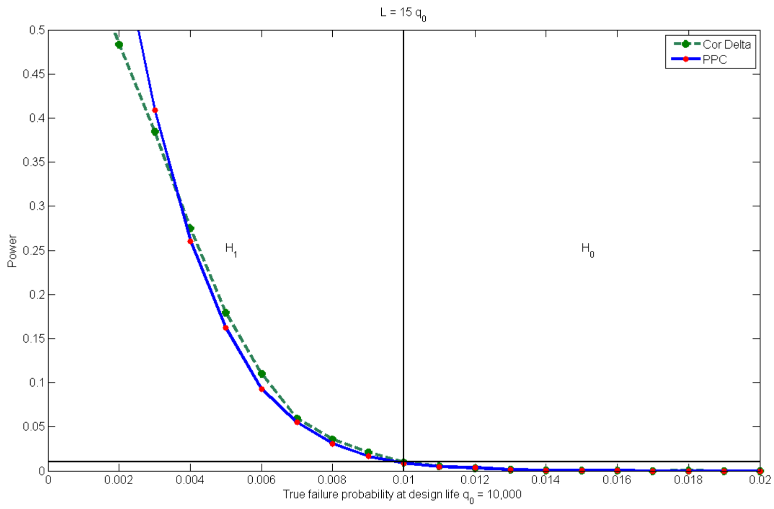

Figure 8.6: Power functions of corrected delta and PPC count for $L = 15q_0$.

8.6 Summary

We finally suggest CUS sampling being method to be preferred. It offers the following advantages:

- Desired significance is achieved.
- Power almost identical to corrected quantile estimates and randomised success runs.
- Free of randomised decisions and ...
- ... need for Monte Carlo simulations, ...
- ... test statistic and critical value allow deterministic calculation.
- Direct interpretation as a generalisation of success runs.
- No discontinuities like in the quantile-CI histograms.
- Robustness as good as for other methods.

Appendix

Appendix A

Lifetime Distributions

When dealing with survival data we first have to restrict to distributions on $[0, \infty)$. We will mainly use two important distributions: Weibull and lognormal. For further classification the notion of the hazard rate is needed.

Definition 13 (Hazard):
Let T be a nonnegative random variable with distribution function (cumulative distribution function, CDF) F. The hazard function is defined to be:

$$h(t) = \lim_{\Delta \to 0+} \frac{\mathbb{P}(t \leq T < t + \Delta \mid t \leq T)}{\Delta}$$

(if the limit exists). For all distributions considered here, the hazard can be written as:

$$h(t) = \frac{f(t)}{1 - F(t)},$$

where f is the density (probability density function, PDF) of F.

The hazard controls the age-specific failure rate, i.e. $h(t)$ is proportional to the probability of failing in $[t, t + \Delta)$, giving time t was survived. Depending on the hazards behaviour, the failure mechanism can be classified:

- **Increasing:** Wearout failures (typical for fatigue applications).
- **Decreasing:** Initial failures (describes infant mortality, e.g. early failures due to manufacturing errors).
- **Constant:** Chance failures (failure due to incorrect operation happening at any time).

For further study see [20] or [22].

A.1 Lognormal

A.1.1 Definition

It was soon recognised, that the well known normal distribution was not suitable to model lifetimes X. The normal fit does however work better for $\ln X$, defining the lognormal distribution:

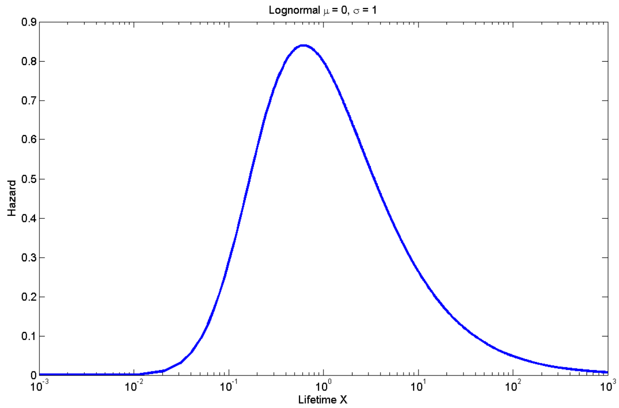

Figure A.1: Hazard of lognormal distribution.

Definition 14 (Lognormal distribution):
A random variable X has a lognormal distribution $\mathcal{LN}(\mu;\sigma^2)$, if and only if $\ln X$ has normal distribution $\mathcal{N}(\mu;\sigma^2)$.

As a direct consequence CDF and PDF are:

$$\begin{aligned} F_{\mu,\sigma^2}(x) &= \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \\ f_{\mu,\sigma^2}(x) &= \frac{1}{\sigma \cdot x}\phi\left(\frac{\ln x - \mu}{\sigma}\right) \end{aligned}$$

(Φ: Standardnomal CDF, ϕ: Standard normal PDF). Quantiles can be simply derived from normal quantiles using an exponential transformation:

$$q_p = e^{\mu + z_p \cdot \sigma},\ z_p = \Phi^{-1}(p)$$

A potential source of confusion is that expectation and variance are not transformed in this way, but:

$$\begin{aligned} \mathbb{E}(X) &= e^{\mu + \frac{\sigma^2}{2}} \\ \mathbb{V}\mathrm{ar}(X) &= e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) \end{aligned}$$

The term e^{μ} is in fact the median of X.

A.1.2 Hazard

There is no easy formula for the hazard of a lognormal distribution. Figure A.1 shows the hazard for lognormal distribution with parameter $\mu = 0$, $\sigma = 1$. The lognormal distribution has the artifact, that up to the median fatigue is modeled, from there on infant failure mechanisms are modeled. Therefore, the lognormal's adequacy for lifetimes is disputable (see also section 3.4.1).

A.2 Weibull

A.2.1 Definition

Based on exponential distributions, Weibull distributions as introduced in [5] are capable of modeling different failure mechanisms.

Definition 15 (Weibull distribution):
A random variable X is called Weibull distributed with parameters $\beta > 0, \gamma > 0$, $\mathcal{W}(\beta;\gamma)$, if $\left(\frac{X}{\beta}\right)^\gamma$ is exponential distributed with $\lambda = 1$.

It follows for CDF and PDF:

$$\begin{aligned} F_{\beta,\gamma}(x) &= 1 - e^{-\left(\frac{x}{\beta}\right)^\gamma} \\ f_{\beta,\gamma}(x) &= \frac{\gamma}{\beta}\left(\frac{x}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x}{\beta}\right)^\gamma} \end{aligned}$$

Quantiles can easily be solved from $F(q_p) = p$ giving:

$$q_p = \beta \cdot (-\ln(1-p))^{\frac{1}{\gamma}}$$

Mean and variance are given by:

$$\begin{aligned} \mathbb{E}(X) &= \beta \cdot \Gamma\left(1 + \frac{1}{\gamma}\right) \\ \mathbb{V}\mathrm{ar}(X) &= \beta^2 \left(\Gamma\left(1 + \frac{2}{\gamma}\right) - \Gamma^2\left(1 + \frac{1}{\gamma}\right)\right) \end{aligned}$$

Here Γ denotes the **gamma function**:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \tag{A.1}$$

with $\Gamma(n) = (n-1)!$ for every $n \in \mathbb{N}$.

A.2.2 Hazard

The calculation of hazards is explicitly possible:

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{\gamma}{\beta}\left(\frac{x}{\beta}\right)^{\gamma-1}$$

It can be seen, that h is a parabola of degree $\gamma - 1$ w.r.t. x. Consequently the sign of $\gamma - 1$ determines whether h is decreasing or increasing:

- $\gamma < 1$: h is decreasing, infant failures are modeled.
- $\gamma = 1$: h is constant, purely random failures are modeled.
- $\gamma > 1$: h is increasing, fatigue life is modeled.

Hence, Weibull distributions can be used to model most common failure mechanisms. Furthermore is has no strange effects like the lognormal hazard maximum. **It is advised to favour Weibull over lognormal**. Therefore, most discussed examples in this thesis will use Weibull distributions.

A.2.3 Relation between quantiles

Consider two quantiles q_{p_1} and q_{p_2}. In fatigue applications we frequently have to switch between failure quotas belonging to different points in time (design life and test duration). It is therefore useful to study the relation between Weibull quantiles:

$$\begin{aligned} \lambda &= \frac{q_{p_2}}{q_{p_1}} = \frac{\beta \cdot (-\ln(1-p_2))^{\frac{1}{\gamma}}}{\beta \cdot (-\ln(1-p_1))^{\frac{1}{\gamma}}} \\ &= \left(\frac{\ln(1-p_1)}{\ln(1-p_2)}\right)^{\frac{1}{\gamma}} \\ \Leftrightarrow (1-p_2) &= (1-p_1)^{\lambda^\gamma} \end{aligned}$$

Using the term **reliability** as the probability of surviving a certain time point, we have:

$$R_2 = R_1^{\lambda^\gamma}$$

This property helped the Weibull distribution to achieve its importance in reliability demonstration, creating the notion **Weibull analysis** (see section 3.4.2).

A.3 Gumbel Distribution

The definition of the Gumbel distribution is not consistent in literature. We propose a new parameterisation in order to have the same relation between Weibull and Gumbel as for lognormal and normal.

Definition 16:
A random variable X is called Gumbel distributed $\mathcal{G}(a;b)$ if it has CDF:

$$F(x) = 1 - e^{-e^{\frac{x-a}{b}}}$$

The density function is:

$$f(x) = \frac{1}{b} e^{\frac{x-a}{b}} e^{-e^{\frac{x-a}{b}}}$$

Quantiles are given by:

$$q_p = a + b\ln(-\ln(1-p))$$

Corollary 2 *If X is Weibull distributed $\mathcal{W}(\beta;\gamma)$, then $\ln X$ is Gumbel distributed $\mathcal{G}\big(\ln\beta;\gamma^{-1}\big)$.*

A Gumbel PDF with $a = 0$, $b = 1$ (standard Gumbel) has its mode at 0, but is not symmetric. Nevertheless, Gumbel PDF are closer to normal distributions than Weibull distributions (see figure A.2). This motivates methods like in appendix B.2.4.

Mean and variance are given by:

$$\begin{aligned} \mathbb{E}(X) &= a - \gamma b \\ \mathbb{V}\mathrm{ar}(X) &= \frac{\pi^2}{6} b^2 \end{aligned}$$

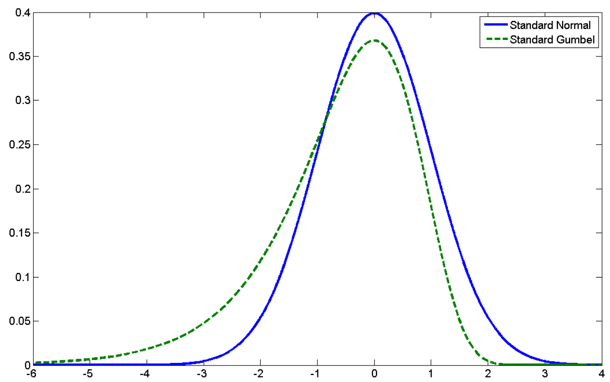

Figure A.2: Standard normal and standard Gumbel PDF.

$\gamma \approx 0.577$ is the Euler-Mascheroni constant. The formulas indicate, that the Gumbel distribution is a scale-location family.

Remark 19:
Fitting a Weibull distribution $\mathcal{W}(\beta; \gamma)$ to data $\underline{T}$ is sometimes done by fitting a Gumbel distribution $\mathcal{G}(a; b)$ to $\underline{\ln T}$. Based on lemma 2, the resulting parameter estimates are:

$$\hat{\beta} = e^{\hat{a}}, \ \hat{\gamma} = \frac{1}{\hat{b}}$$

The required likelihood can be found in section B.2.1.

A.4 Beta Distribution

A.4.1 Definition

The beta distribution is not a lifetime distribution, but can be used to model prior information on a components reliability (as is done from section 7.1.2 on).

Definition 17 (Beta distribution):
A random variable with density

$$f(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}$$

is called beta distributed, Beta$(p; q)$. *Here*

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^1 x^{p-1}(1-x)^{q-1}dx, \qquad \text{(A.2)}$$

where Γ *is the gamma function from equation (A.1).*

For integer p, q there is an explicit formula for the CDF F:

Lemma 8 *For $p, q \in \mathbb{N}$ the CDF of a* Beta$(p; q)$ *distribution is given by:*

$$F_{p,q}(x) = \sum_{j=p}^{p+q-1} \frac{(p+q-1)!}{j!(p+q-1-j)!} x^j (1-x)^{p+q-1-j} \tag{A.3}$$

Proof *Induction on q.*
If $q = 1$ then:

$$F_{p,1}(x) = \int_0^x \frac{1}{B(p,1)} t^{p-1} dt = \frac{x^p}{pB(p,1)} = x^p$$

For general q perform integration by parts:

$$\begin{aligned}
F_{p,q}(x) &= \int_0^x \frac{1}{B(p,q)} t^{p-1}(1-t)^{q-1} dt \\
&= \frac{1}{B(p,q)} \left(\left[\frac{t^p}{p}(1-t)^{q-1} \right]_0^x + \frac{q-1}{q} \int_0^x t^p (1-t)^{(q-1)-1} \right) \\
&= \frac{(p+q-1)!}{p!(q-1)!} x^p (1-x)^{q-1} + \frac{q-1}{q} \cdot \frac{B(p+1,q-1)}{B(p,q)} F_{p+1,q-1}(x) \\
&= \frac{(p+q-1)!}{p!(q-1)!} x^p (1-x)^{q-1} + \sum_{j=p+1}^{p+q-1} \frac{(p+q-1)!}{j!(p+q-1-j)!} x^j (1-x)^{p+q-1-j} \\
&= \sum_{j=p}^{p+q-1} \frac{(p+q-1)!}{j!(p+q-1-j)!} x^j (1-x)^{p+q-1-j}
\end{aligned}$$

qed

Mean and variance are given by:

$$\begin{aligned}
\mathbb{E}(X) &= \frac{p}{p+q} \\
\mathbb{V}\mathrm{ar}(X) &= \frac{pq}{(p+q+1)(p+q)^2}
\end{aligned}$$

A.4.2 Beta Priors

As already mentioned, beta distributions are needed to model prior information on a components reliability R. Figure A.3 shows different possible PDF. It can be seen, that the case $p > q$ puts mass on values close to 1, while $p < q$ does the opposite.

In this setup there is possible reparameterisation of the beta distribution using $N = p + q$ instead of q as second parameter. For reasons that will become clear in chapter 7.1.2 the parameters are also called **pseudo number of survivors** p, **pseudo number of survivors** q and pseudo sample size N.

Remark 20 (Uniform priors):
A special case of the beta distribution is Beta$(1; 1)$, which defines a uniform distribution on $[0, 1]$. This will be used in section 7.1.1.

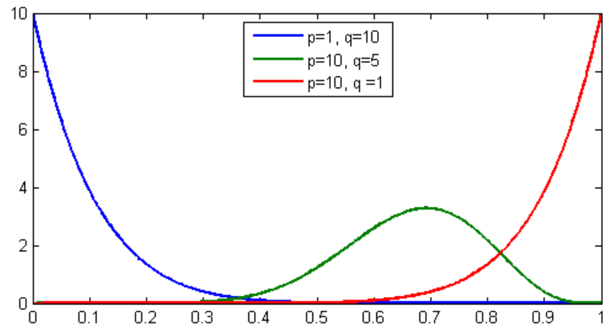

Figure A.3: Density function for different values of p, q.

Appendix B

Delta Method for Weibull Distributions

When applying the delta method to a distribution, all derivatives of second order have to be known. This chapter provides the necessary formulas.

B.1 Weibull

B.1.1 Likelihood

The PDF and CDF of a Weibull distribution are:

$$\begin{aligned} f(x) &= \frac{\gamma}{\beta}\left(\frac{x}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x}{\beta}\right)^\gamma} \\ F(x) &= 1 - e^{-\left(\frac{x}{\beta}\right)^\gamma} \end{aligned}$$

Therefore, a sample $\underline{x}$ with censoring indicator $\underline{\delta}$ results in the likelihood:

$$\begin{aligned} L(\beta, \gamma \mid \underline{x}) &= \left(\frac{\gamma}{\beta}\right)^{N_u} \cdot \left(\frac{\prod_{i:\,\delta_i=1} x_i}{\beta^{N_u}}\right)^{\gamma-1} \cdot e^{-\sum_{i=1}^{N}\left(\frac{x_i}{\beta}\right)^\gamma} \\ l(\beta, \gamma \mid \underline{x}) &= N_u \ln\gamma - N_u \ln\beta + (\gamma - 1)\sum_{i:\,\delta_i=1} \ln\frac{x_i}{\beta} - \sum_{i=1}^{N}\left(\frac{x_i}{\beta}\right)^\gamma \end{aligned}$$

where $N_u = \sum \delta_i$ is the number of uncensored data in the sample.

B.1.2 Score and Fisher Information

$$\begin{aligned} \frac{\partial}{\partial\beta} l(\beta, \gamma \mid \underline{x}) &= -\frac{N_u\gamma}{\beta} + \frac{\gamma}{\beta^{\gamma+1}}\sum_{i=1}^{N} x_i^\gamma \\ \frac{\partial^2}{\partial\beta^2} l(\beta, \gamma \mid \underline{x}) &= \frac{N_u\gamma}{\beta^2} - \frac{\gamma(\gamma+1)}{\beta^{\gamma+2}}\sum_{i=1}^{N} x_i^\gamma \end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial\gamma}l(\beta,\gamma\mid\underline{x}) &= \frac{N_u}{\gamma}+\sum_{i:\,\delta_i=1}\ln x_i - N_u\ln\beta-\sum_{i=1}^{N}\left(\frac{x_i}{\beta}\right)^{\gamma}\cdot\ln\left(\frac{x_i}{\beta}\right)\\
\frac{\partial^2}{\partial\gamma^2}l(\beta,\gamma\mid\underline{x}) &= -\frac{N_u}{\gamma^2}-\sum_{i=1}^{N}\left(\frac{x_i}{\beta}\right)^{\gamma}\cdot\ln^2\left(\frac{x_i}{\beta}\right)\\
\frac{\partial^2}{\partial\gamma\partial\beta}l(\beta,\gamma\mid\underline{x}) &= -\frac{N_u}{\beta}+\frac{\gamma}{\beta}\sum_{i=1}^{N}\left(\frac{x_i}{\beta}\right)^{\gamma}\cdot\ln\left(\frac{x_i}{\beta}\right)+\frac{1}{\beta}\sum_{i=1}^{N}\left(\frac{x_i}{\beta}\right)^{\gamma}
\end{aligned}$$

B.1.3 Maximum Likelihood Estimates

From the upper equations the MLE can be written as:

$$\hat{\beta}^{\hat{\gamma}} = \frac{1}{N_u}\sum_{i=1}^{N}X_i^{\hat{\gamma}} \tag{B.1}$$

$$N_u\sum_{i=1}^{N}\left(X_i^{\hat{\gamma}}\cdot\ln X_i\right) = \frac{N_u}{\hat{\gamma}}\sum_{i=1}^{N}X_i^{\hat{\gamma}}+\sum_{\delta_i=1}\ln X_i\cdot\sum_{i=1}^{N}X_i^{\hat{\gamma}} \tag{B.2}$$

The second equation is obtained from plugging in $\frac{\partial}{\partial\beta}l=0$ into $\frac{\partial}{\partial\gamma}l=0$. While the first equation delivers estimates $\hat{\beta}$ if $\hat{\gamma}$ is known, $\hat{\gamma}$ is the numerical solution of the second equation.

B.1.4 Estimated Fisher Information

Plugging in the MLE into the second derivatives gives:

$$\begin{aligned}
\frac{\partial^2}{\partial\beta^2}l(\hat{\beta},\hat{\gamma}\mid\underline{x}) &= -\frac{N_u\hat{\gamma}^2}{\hat{\beta}^2}\\
\frac{\partial^2}{\partial\gamma^2}l(\hat{\beta},\hat{\gamma}\mid\underline{x}) &= -\frac{N_u}{\hat{\gamma}^2}-\sum_{i=1}^{N}\left(\frac{x_i}{\hat{\beta}}\right)^{\hat{\gamma}}\ln\frac{x_i^2}{\hat{\beta}^2}\\
\frac{\partial^2}{\partial\gamma\partial\beta}l(\hat{\beta},\hat{\gamma}\mid\underline{x}) &= \frac{N_u}{\hat{\beta}}-\frac{N_u\hat{\gamma}}{\hat{\beta}}+\frac{\hat{\gamma}}{\hat{\beta}}\sum_{i=1}^{N_u}\ln x_i
\end{aligned}$$

B.2 Gumbel

B.2.1 Likelihood

The density and distribution function of a Gumbel distribution have the form:

$$\begin{aligned}
f(x) &= \frac{1}{b}e^{\frac{x-a}{b}}e^{-e^{\frac{x-a}{b}}}\\
F(x) &= 1-e^{-e^{\frac{x-a}{b}}}
\end{aligned}$$

Therefore a sample $\underline{x}$ with censoring indicator $\underline{\delta}$ gives the likelihood:

$$
\begin{aligned}
L(a,b \mid \underline{x}) &= \prod_{\delta_i=1} \frac{1}{b} e^{\frac{x_i-a}{b}} e^{-e^{\frac{x_i-a}{b}}} \prod_{\delta_i=0} e^{-e^{\frac{x_i-a}{b}}} \\
l(a,b \mid \underline{x}) &= -N_u \ln b + \frac{\sum_{\delta_i=1} x_i - N_u a}{b} - \sum_{i=1}^{N} e^{\frac{x_i-a}{b}}
\end{aligned}
$$

where $N_u = \sum \delta$ is the number of uncensored data in the sample.

B.2.2 Score and Fisher information

$$
\begin{aligned}
\frac{\partial}{\partial a} l(\beta,\gamma \mid \underline{x}) &= -\frac{N_u}{b} + \frac{1}{b}\sum_{i=1}^{N} e^{\frac{x_i-a}{b}} \\
\frac{\partial^2}{\partial a^2} l(\beta,\gamma \mid \underline{x}) &= -\frac{1}{b^2}\sum_{i=1}^{N} e^{\frac{x_i-a}{b}} \\
\frac{\partial}{\partial b} l(\beta,\gamma \mid \underline{x}) &= -\frac{N_u}{b} - \frac{\sum_{\delta_i=1} x_i - N_u a}{b^2} + \sum_{i=1}^{N} \frac{x_i-a}{b^2} e^{\frac{x_i-a}{b}} \\
\frac{\partial^2}{\partial b^2} l(\beta,\gamma \mid \underline{x}) &= \frac{N_u}{b^2} + 2\frac{\sum_{\delta_i=1} x_i - N_u a}{b^3} \sum_{i=1}^{N} \left(-\left(\frac{x_i-a}{b^2}\right)^2 - 2\frac{x_i-a}{b^3} \right) e^{\frac{x_i-a}{b}} \\
\frac{\partial^2}{\partial b \partial a} l(\beta,\gamma \mid \underline{x}) &= \frac{N_u}{b^2} - \frac{1}{b^2}\sum_{i=1}^{N} e^{\frac{x_i-a}{b}} - \frac{1}{b^3}\sum_{i=1}^{N} (x_i-a) e^{\frac{x_i-a}{b}}
\end{aligned}
$$

B.2.3 Maximum Likelihood Estimates

From the upper equations the MLE can be written as solution of:

$$
\begin{aligned}
N_u &= \sum_{i=1}^{N} e^{\frac{x_i-a}{b}} \\
N_u a &= \sum_{\delta_i=1} x_i + N_u b - \sum_{i=1}^{N} e^{\frac{x_i-a}{b}} (x_i - a)
\end{aligned}
$$

Numerical solving gives estimates $\hat{a}, \hat{b}$.

B.2.4 Logarithmic Delta Method

Assume that a quantile q_p has to be estimated in a Weibull model $\mathcal{W}(\beta;\gamma)$. To ensure positive CI, logarithms shall be used, and a Gumbel model $\mathcal{G}(a;b)$ is fitted to the data $\underline{\ln T}$ (using remark 19). Pluging in in the ML estimates for a, b into the likelihood's Hessian gives the Hesse estimator $\hat{I}$ of definition 3. A quantile CI for $\ln q_p$ is obtained as:

$$
\widehat{\ln q}_{p,\alpha} = \widehat{\ln q_p} + t_{N-1}^{-1}(1-\alpha) \cdot \hat{\sigma}_\beta
$$

The quantile functions estimated variance $\hat{\sigma}_\beta$ is obtained from Gaussian error propagation for $(\ln q_p)(a, b) = a + b \cdot g_p$, where g_p is the standard Gumbel quantile.

Transforming the quantile CI back to the scale of $\underline{T}$ gives the **logarithmic delta method**:

$$\hat{q}_{p,\alpha} = e^{\widehat{\ln q_p}_{,\alpha}} = \hat{q}_p \cdot \exp\left(t_{N-1}^{-1}(1-\alpha) \cdot \hat{\sigma}_\beta\right)$$

Appendix C

Proof of the CUS Density

C.1 Proving the General Formula

Let G_k, G_k^+ denote the CDFs of g_k, g_k^+ and set $F = F_{\theta_0}(L)$, $\bar{F} = 1 - F_{\theta_0}(L)$ as short forms.

Induction Start Given $T \leq L$, the random variables T, T^+ can not be distinguished. In case of $T > L$ the CDF G_1^+ has to collect all mass of G_1 in the single point F (see figure C.1). It follows:

$$G_1^+(u) = G_1(\min(u, L)) + \mathbb{1}_{[F,1]}(u)$$

The derivative of G_1^+ gives the claimed density :

$$g_1^+(u) = g_1\left(\frac{u}{F}\right) + \bar{F}\delta_F(u) = \mathbb{1}_{[1,F]}(u) + \bar{F} \cdot \delta_F(u)$$

Induction Step

$$\begin{aligned}
&\int_{-\infty}^{\infty} \Big(\sum_{i=0}^{N-1} \binom{N}{i} F^{N-1-i} \bar{F}^i g_{N-i}\left(\frac{u}{F} - i\right) + \bar{F}^N \delta_{NF}(u) \Big) \\
&\quad + \big(\mathbb{1}_{[0,F]}(u-s) + \bar{F}\delta_F(u-s) \big)\, ds \\
= &\quad \sum_{i=0}^{N-1} \binom{N}{i} F^{N-1-i} \bar{F}^i \int_{-\infty}^{\infty} \mathbb{1}_{[0,F]}(u-s)\, g_{N-i}\left(\frac{s}{F} - i\right) ds \\
&\quad + \sum_{i=0}^{N-1} \binom{N}{i} F^{N-1-i} \bar{F}^{i+1} \int_{-\infty}^{\infty} \mathbb{1}_{[0,F]}(u-s)\, g_{N-i}\left(\frac{s}{F} - i\right) \\
&\quad \cdot \delta_F(u-s) ds + \bar{F}^N \int_{-\infty}^{\infty} \mathbb{1}_{[0,F]}(u-s)\, \delta_{NF}(s) ds \\
&\quad + \bar{F}^{N+1} \int_{-\infty}^{\infty} \delta_{NF}(s) \delta_F(u-s) ds
\end{aligned}$$

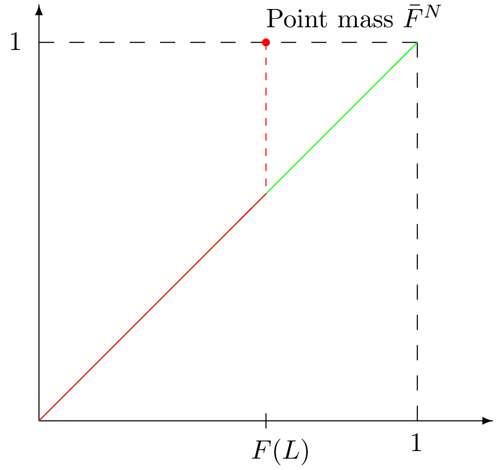

Figure C.1: Distribution function of $F(T^+)$ (red) and $F(T)$ (green).

$$
\begin{aligned}
&= \sum_{i=0}^{N-1} \binom{N}{i} F^{N-i}\bar{F}^i g_{N+1-i}\left(\frac{u}{F}-i\right) \\
&\quad + \sum_{j=0}^{N-1} \binom{N}{j} F^{N-1-j}\bar{F}^{j+1} g_{N-j}\left(\frac{u}{F}-(j+1)\right) \\
&\quad + \bar{F} g_1\left(\frac{u}{F}-N\right) + \bar{F}^{N+1}\delta_{(N+1)F}(u) \\
&= \sum_{i=0}^{N-1} \binom{N}{i} F^{N-i}\bar{F}^i g_{N+1-i}\left(\frac{u}{F}-i\right) \\
&\quad + \sum_{i=1}^{N} \binom{N}{i-1} F^{N-i}\bar{F}^i g_{N+1-i}\left(\frac{u}{F}-i\right) \\
&\quad + \bar{F} g_1\left(\frac{u}{F}-N\right) + \bar{F}^{N+1}\delta_{(N+1)F)}(u) \\
&= \sum_{i=0}^{N} \binom{N+1}{i} F^{N-i}\bar{F}^i g_{N+1-i}\left(\frac{u}{F}-i\right) + \bar{F}^{N+1}\delta_{(N+1)F}(u)
\end{aligned}
$$

C.2 Formulas for Small N

In section 5.4.2 it was shown, that a CUS($N,1$) distribution may be approximated by a normal distribution for $N \geq 5$. For smaller N the exact PDF has to be calculated convoluting a uniform distribution on $[0,1]$. The resulting PDFs are:

$$
\begin{aligned}
g_1(x) &= \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{else} \end{cases} \\
g_2(x) &= \begin{cases} x, & 0 \leq x < 1 \\ 2-x, & 1 \leq x \leq 2 \\ 0, & \text{else} \end{cases} \\
g_3(x) &= \begin{cases} \frac{x^2}{2}, & 0 \leq x < 1 \\ -x^2+3x-\frac{3}{2}, & 1 \leq x < 2 \\ \frac{(x-1)^2}{2}-2x+4, & 2 \leq x \leq 3 \\ 0, & \text{else} \end{cases} \\
g_4(x) &= \begin{cases} \frac{x^3}{6}, & 0 \leq x < 1 \\ -\frac{3x^3-12x^2+12x-4}{6}, & 1 \leq x < 2 \\ \frac{x^3}{2}-4x^2-10x-\frac{22}{3}, & 2 \leq x < 3 \\ -\frac{x^3-12x^2+48x-64}{6}, & 3 \leq x \leq 3 \\ 0, & \text{else} \end{cases}
\end{aligned}
$$

Appendix D

Introduction to Bayesian Statistics

Accounting for previous information in statistical inference is done by using Bayesian statistics. For reliability analysis this would allow to use knowledge about the quality of similar units, which are already observed over a longer period. To get simple *knowledge update* formulas, the concept of conjugated families plays a key role.

D.1 Conjugated Families

D.1.1 General Definition

Let $\theta \in \Theta \subset \mathbb{R}^k$ be the unknown parameter of a distribution family

$$\mathcal{F} = \{F_\theta \mid \theta \in \Theta\}, \text{ with } X_1, \ldots, X_N \overset{\text{iid}}{\sim} F_\theta$$

Previous information about θ is formulated by means of the prior density $\pi(\theta)$. E.g., using a normal PDF π with mean μ and variance σ^2 implies θ to lie in the interval $\mu \pm 2\sigma$ with plausibility 95.45%.

After collecting data $\underline{X} = \underline{x}$ Bayes' theorem allows updating the **prior density** π to a **posterior density** $\pi(\cdot \mid \underline{x})$ via:

$$\pi(\theta \mid \underline{x}) = \frac{\pi(\theta) \cdot f_{\underline{X}}(\underline{x} \mid \theta)}{\int\limits_{\Theta} \pi(\theta) \cdot f_{\underline{X}}(\underline{x} \mid \theta) d\theta} \tag{D.1}$$

In general $\pi(\theta \mid \underline{x})$ will be from a different distribution family than $\pi(\theta)$, which might not even be of a common form.

Definition 18 (Conjugated family):
A family Π of prior distributions is said to be conjugated to a family $\mathcal{F}$ of sampling distributions, if for every true sampling distribution $F \in \mathcal{F}$, every sample size $N \in \mathbb{N}$ and every realisation $\underline{x}$ the posterior distribution $\pi(\cdot \mid \underline{x})$ belongs to Π.

Using conjugated families often allows for simple formulas for getting the posterior PDF parameters like in equation 7.6 or lemma 6. In this case the integral in the denominator of equation (D.1) does not have to be calculated numerically.

Example 21 (Conjugated family of gamma distributions):
Let X be Poisson distributed with unknown parameter μ. If the prior density is given by a gamma distribution $\Gamma(\nu, \lambda)$, then the posterior density is also a gamma distribution $\Gamma(\nu + \sum x_i, \lambda + N)$. Hence one has the parameter update formula $(\bar{\nu}, \bar{\lambda}) = (\nu + \sum x_i, \lambda + N)$.

D.2 Bayesian Confidence Intervals

As soon as the posterior $\bar{\pi}$ is available, point and interval estimates might be derived. There are two possibilities for defining point estimates:

- $\hat{\theta} = \mathbb{E}(\bar{\pi})$: If θ is interpreted as a random variable, then in terms of mean squared error the posterior expectation is the best guess for the unknown parameter.
- $\hat{\theta} = \text{argmax}\,\bar{\pi}(\theta)$: Generalises the likelihood approach, by considering the posterior as a likelihood merged with prior information.

E.g. the beta distribution $\text{Beta}(p; q)$ has mode $\frac{p-1}{p+q-2}$ and expectation $\frac{p}{p+q}$. In connection to the theory of section 7.2 p was called pseudo number of survivors and $p+q$ the pseudo sample size and θ was the a components reliability. This gives the expectation a clear interpretation, while it is not intuitive why the pseudo number of survivors should be reduced by one in the mode.

Defining confidence intervals for θ is done with the notion of **highest (posterior) density regions**:

Definition 19 (Highest density region HDR):
An interval $I \subset \Theta$ is called highest (posterior) density region if $\bar{\pi}(\theta_1) > \bar{\pi}(\theta_2)$ for each $\theta_1 \in I$, $\theta_2 \in \Theta \setminus I$.

D.3 Sequential Knowledge Update

Applying Bayes' Theorem allows for knowledge update after each sample. There is an important property making this update consistent:

Lemma 9 *Consider two independent samples $\underline{X}_1$, $\underline{X}_2$ available for inference about a parameter θ. Let $\underline{X} = (\underline{X}_1, \underline{X}_2)$ denote the combined sample, then:*

$$\pi(\theta \mid \underline{X}) \propto \pi(\theta \mid \underline{X}_1) L(\theta \mid \underline{X}_2)$$

Proof *Since $\underline{X}_1$, $\underline{X}_2$ are independent, it holds that $L(\theta \mid \underline{X}) = L(\theta \mid \underline{X}_1) \cdot L(\theta \mid \underline{X}_2)$. It follows directly:*

$$\begin{aligned}
\pi(\theta \mid \underline{X}) &\propto \pi(\theta) L(\theta \mid \underline{X}) \\
&\propto \pi(\theta) L(\theta \mid \underline{X}_1) L(\theta \mid \underline{X}_2) \\
&\propto \pi(\theta \mid \underline{X}_1) L(\theta \mid \underline{X}_2)
\end{aligned}$$

qed

The lemma can even be used to update the prior after each observation. This will be used in section 7.2.2, when the posterior is calculated w.l.o.g. for the case $N = 1$ for reasons of simplicity.

List of Figures

Bibliography

[1] "*Betriebsfestigkeit: Verfahren und Daten zur Bauteilberechnung*", Erwin Haibach, Springer 2006 Berlin, Heidelberg.

[2] "*Zuverlässigkeit im Fahrzeug- und Maschinenbau*", Benrd Bertsche, Gisbert Lechner, Springer, 2004 Berlin.

[3] "*Qualitätsmanagement in der Automobilindustrie 3, Zuerlässigkeitssicherung bei Automobilherstellern und Lieferanten Teil2*", Verband der Automobilindustrie VDA, 3. überarbeitete und erweiterte Auflage 2000.

[4] "*Bayesian techniques to reduce sample size in automotive electronics attribute testing*", Kleyner, Bhagath, Gasparini, Robinson, Bender. Microelectronic Reliability, vol 37, No 6, pp 879-883, 1997.

[5] "*A Statistical Theory of the Strength of Materials*, Waloddi Weibull, Ingeniörsvetenskapsakademiens Handlingar Nr 151, 1939.

[6] "*The Statistical Treatment of Fatigue Experiments*, Leonard G. Johnson, Research Laboratories General Motors Corporation, 1959.

[7] "*Planung von Zuverlässigkeitstests mit weitreichender Berücksichtigung von Vorkenntnissen*", Anna Krolo, Berichte aus dem Institut für Maschinenelemente, Universität Stuttgart, 2004.

[8] "*Considering Prior Information for Accelearted Tests with a Lifetime-Ratio*", Anna Krolo, Bettina Rzepka, Bernd Bertsche, Proceedings of 3rd International Conference on Mathematical Methods in Reliability, Methodology and Practice, June 17-20, 2002, Trondheim Norway.

[9] "*Statistik*", Joachim Hartung, Oldenburg Verlag München, 2005.

[10] "*Statistical Inference*", George Casella, Roger L. Berger, Duxbury Advanced Series, 2002 Pacific Grove, Calif..

[11] "*Asymptotic Statistics*", A.W. van der Vaart, Cambridge University Press, 1998.

[12] "*Significance Tests for Patterns in Continuous Data*", R.J. Bolton, D.J. Hand, Dept. of Math., Imperial Coll. London, Data Mining, 2001, Proceedings IEEE.

[13] "*Approximation theorems of mathematical statistics*", Robert J. Serfling, Wiley Series in probability and mathematical statistics, 1980 New York.

[14] "*Robust Statistics*", Peter J. Huber, Wiley Series in probability and mathematical statistics, 1981 New York.

[15] "*The jackknife, the bootstrap and other resampling plans*", Bradley Efron, Regional conference series in applied mathematics, 1982 Philadelphia.

[16] "*The Jackknife and Bootstrap*", Jun Shao, Dongsheng Tu, Springer-Verlag, 1996.

[17] "*The Bootstrap and Egdeworth Expansion*", Peter Hall, Springer-Verlag, 1992 New York.

[18] "*Resampling-Methoden zur mse-Korrektur und Anwendungen in der Betriebsfestigkeit*", Sascha Feth, Jürgen Franke, Michael Speckert, Berichte des Fraunhofer ITWM, Nr. 116 (2007).

[19] "*Empirical likelihood and small samples*", Art Owen Computing science and statistics, 1992, pp. 69-88.

[20] "*Analysis of survival data*", D.R. Cox, D. Oakes, Chapman and Hall, 1984 London.

[21] "*Bis das Getriebe kracht: Statistik ersetzt teure Tests*", H. Dette, Chr. I. Kozub, Rubin Wissenschaftsmagazin 1/00 Ausgabe 1/2000.

[22] "*Life Testing and Reliability Estimation*", S.K. Sinha, B.K. Kale, Wiley 1980 New Delhi.

[23] "*Right Censoring and the Kaplan-Meier and Nelson-Aalen Estimators*", Martin Jacobsen, Institute of Mathematical Statistics, University of Copenhagen, Prepring 1986,6 Copenhagen.

[24] "*Bayesian Theory*", Jose M. Bernado, Adrian F. M. Smith, Wiley Series in Probability and mathematical statistics, 1993.

[25] "*Bayesian Inference in Statistical Analysis*", George E.P. Box, George C. Tiao, Addison-Wesley Series in Behavioral Science: Quantitative Methods, 1973.

[26] "*Bayesian Reliability Analysis*", Harry F. Martz, Ray A. Waller, Wiley Series in Probability and Mathematical Statistics, 1982 New York.

[27] "*Bayesian Statistics: An Introduction*"', Peter M. Lee, Oxford University Press, New York, 1989.

Index

Curriculum Vitae

Name: Sascha Feth (birth name Klos)

Date of birth: March 9, 1982 in Landstuhl, Germany.

Education

1988 - 1992 Primary shool Wendelinus-Grundschule Ramstein-Miesenbach.

1992 - 1999 Secondary shool Konrad-Adenauer-Realschule Landstuhl.

1999 - 2002 Grammar shool Sickingen-Gymnasium Landstuhl.

2002 - 2006 Mathematical studies at the university of technology Kaiserslautern.

2006 - 2009 Phd studies in mathematic at the university of technology Kaiserslautern.