A statistical question motivated by testing of quantum devices

Berlin, Marieke K.

Award date:
2021

Department of Mathematics and Computer Science

A statistical question motivated by testing of quantum devices

Bachelor final project

M.K. Berlin
Student number:
1252097

Supervisor:
Dr. R.M. Pires Da Silva Castro

Eindhoven, January 2021

Abstract

This bachelor final project is motivated by a problem of quantum devices. Some elements (qubits and gates) in the quantum devices are said to be degrading and thus unreliable. Therefore, experiments are done. After some time t, one observes whether such an element is broken or not. The probability that the element is broken is denoted by $f(t)$, which is a function that is assumed to be linear for small values of t. Two problems will be considered. Can we test if the function f is linear in the range of times we observed? And can we estimate the derivative of f at 0, with and without making further assumptions on this function? To test if the function f is linear, a bootstrap method is used. The quality of this bootstrap method is analyzed using simulations. An estimator of the derivative of f at 0 is determined, under the assumption that f is linear. Finally, the intervals method is defined. This method gives an estimate of the derivative of $f(t)$ at 0 without making assumptions on f. This method is based upon a sequence of estimates and its confidence intervals. It selects the estimate which uses as much as possible observations and at the same time is in all confidence intervals of previous estimates (which use less observations). Overall, this report gives some insight in how the behavior of the degrading elements can be investigated.

Contents

Chapter 1

Introduction and problem definition

1.1 Introduction

Quantum computers and related quantum devices are believed to be able to address scientific challenges that current technology can not tackle at this moment. Some examples of these problems can be found in the article written by Loeffler [4]. One of the application areas is the financial services. Due to the computational power of a quantum computer, significantly more variables can be added to the existing models. This will make these models more precise. Other problems that will be solved by a quantum device are optimization problems. Solving the optimization problems will help in many application areas such as supply chain logistics. Furthermore, according to the report [5] the main motivation behind developing a quantum computer is solving Shor's algorithm which gives the prime factors of an integer.

Quantum technology can be categorized into four different application areas. These are 'quantum computing, quantum simulation, quantum communication and quantum sensing'[1]. A brief explanation of each of the four application areas will be given.

A quantum computer has qubits which essentially play the role of bits in a traditional computer. In a traditional computer bits are stored in a simple transistor-based circuit that is in one of two states (0 and 1). A qubit is a much more complex object, that can be in a wide range of possible states. According to Sara Gamble, one of the authors of [5], this will result in a revolution in computation and the solving of certain types of intractable problems.

A quantum simulator is a small quantum computer which is programmed to simulate one specific process. Therefore, it will only be able to solve a limited number of problems [1].

A quantum communication device can replace all kinds of communication tools. These are often secured with RSA, which is an encryption that can easily be cracked by a quantum device. The only known solution which can make communication safe after a quantum device is introduced is to send the encryption code via a quantum communication device [1].

Finally, a quantum sensing device could detect and react to much more than existing sensing devices. The quantum sensing devices will likely lead to more powerful and sensitive measuring technology[1].

It may be clear that quantum technology will change not only the field of computing as mentioned before, but all four application areas.

The main motivation behind this bachelor final project is the fact that qubits and gates (which are any operations that change the state of a qubit) are currently not stable enough to perform accurately for a substantial amount of time. According to de Graaf et all., 'a long period of qubit performance degradation can cause a catastrophic event for quantum computation'[3]. Therefore, reliability is at stake. It's argued that the performance of these parts degrades over time up to some point after which they are completely unpredictable. Our knowledge about the degrading behavior of these elements is limited. Thus, the goal of this project is to understand and estimate some properties of this degrading process.

The above realization sets the stage for this project. Namely we consider a setting where experiments are conducted on a device in order to understand the degradation process. One experiment will result in a failure or a success. This means that the results of the experiments will be binary (0 if the experiment was a success and 1 if the experiment was a failure). Unfortunately, these binary observations do not contain much information, causing a limited knowledge about the failure probability. It is challenging to come up with a simple model of degradation over time, and therefore a non-parametric stance will be taken. This will be explained in more detail in the problem definition, section 1.2. At this moment there is no parametric model for the degradation function.

This report consists of five chapters. In section 1.2 the problem of the degrading elements is defined in a mathematical setting. Chapter 2 introduces an estimator for a specific case and its quality is investigated. Chapter 3 focuses on a test which investigates the linearity of the failure probability. In Chapter 4 an estimator is given for the derivative of the failure probability (b) in a more general setting. Finally, some conclusions and a discussion will be given in Chapter 5.

1.2 Problem definition and notation

The main interest of this bachelor final project lies with the behaviour of degrading elements of a quantum device. To investigate the behaviour of these elements, an experiment is constructed. This experiment starts with an element which is set into a given state. After some time T the state of this element will be measured again. If the element is still in the same state it will be considered a success, otherwise, it will be considered a failure. This experiment will be repeated n times. If the ith experiment, where $i \in \{1, \ldots, n\}$, was successful then this will be denoted by $Y_i = 0$ and $Y_i = 1$ otherwise, where T_i is defined as the time of measurement of the ith experiment. The values of T_i can be assumed to be random or deterministic. For this report it was assumed that T_i is deterministic. The results that will be available after the experiment are given in the form (Y_i, T_i) for $i = 1, \ldots, n$. The probability of a failure (probability of $Y_i = 1$) is believed to increase over time. This implies that for larger values of T_i the probability that the element is in another state increases as well. This failure probability is denoted by the function f where the parameter t will be mapped to $f(t)$ and $f : [0, \infty) \to [0, 1]$. Around the origin this function f is assumed to be linear. Formally, we assume that $f(t) = a + b \cdot t + o(t)$ for $t \to 0$. The values of a and b determine the behaviour of the elements. Notice that $f(0) = a$. We are interested in a case where we know that an element can hold a state for at least a small amount of time. Thus, we consider that $a = 0$. Therefore, the main quantity of interest of this experiment is b.

Once the experiments are done the values of $Y_1, \ldots, Y_n$ are collected which correspond to the times $T_1, \ldots, T_n$. If the values of $T_1, \ldots, T_n$ are given the observations $Y_1, \ldots, Y_n$ are assumed to be independent samples from the model. This model is assumed to be a Bernoulli distribution with parameter $f(t)$. This results into the first model

$$Y_i | T_1, \ldots, T_n \overset{\text{independent}}{\sim} \text{Bernoulli}(f(T_i)). \tag{1.1}$$

In some settings the Bernoulli model leads to a few complications. Therefore, another model, but related one, will also be considered. Note that if the parameter of a Bernoulli distribution is small, this distribution is similar to the Poisson distribution by the Poisson limit theorem[2]. We are interested in the case of a Bernoulli model with a small parameter. This distribution might be more insightful. The Poisson model will be defined as

$$Y_i|T_1,\ldots,T_n \overset{\text{independent}}{\sim} \text{Poisson}(f(T_i)). \tag{1.2}$$

With the results from the experiment and the assumed models, the behaviour of the degrading elements will be investigated.
The goal of this research can be split into two parts. The first is to test if the function $f(t)$, that indicates the failure probability, is truly linear and the second part is to estimate the quantity of interest b. These problems will be solved using three sub-questions.

1. If f is actually linear, how can b be estimated effectively?
2. Is it possible to test if f is linear?
3. How can b be estimated in the general setting of $f(t) = a + b \cdot t + o(t)$?

Each sub-question will be addressed in a distinct chapter.

Chapter 2

Estimating b when f is linear

One of the main goals of this project is to estimate the derivative of the linear function $f(t)$ at zero, which is b. In this first chapter some simplifications will be assumed to be able to estimate this quantity. These are that:

1. the values of $T_1, \dots, T_n$ are deterministic, and given by $t_1, \dots, t_n$
2. $a = 0$ so, $f(t) = b \cdot t$

In section 1.2 two different problem formulations were considered. The first model (1.1) where $Y_i | T_1, \dots, T_n \overset{\text{independent}}{\sim} \text{Bernoulli}(f(T_i))$ is based on Bernoulli observations. Note, that T_i is assumed to be deterministic in this case and therefore the random variables Y_i are mutually independent. Similarly, the second model (1.2) is based on Poisson observations. For both models the method of moments estimator (MME) and the maximum likelihood estimator (MLE) are determined. These are classical estimation methods which can be reviewed in the book by Abramovich [2]. Then the estimator will be assessed with the mean squared error (MSE).

2.1 Estimator of b assuming model 1 (Bernoulli)

The estimators will first be determined for model 1.1. Here

$$Y_i \sim \text{Bernoulli}(f(t_i)) \text{ for } i \in \{1, \dots, n\}.$$

The method of moments will be applied to determine a first estimator of b. Note that Y_i for $\imath \in \{1, \dots, n\}$ is not identically distributed. Therefore, the law of large numbers as stated in theorem 5.2 in [2] cannot be applied. However, like used in the law of large number the quantity $\frac{1}{n}\sum_{i=1}^{n} Y_i$ is investigated. Clearly

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[Y_i\right] = \frac{1}{n} \cdot b \sum_{i=1}^{n} t_i$$

and

$$\text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} f(t_i)(1 - f(t_i)) \leq \frac{1}{n^2}\sum_{i=1}^{n} \frac{1}{4} = \frac{1}{4n}.$$

So, the variance of $\frac{1}{n}\sum_{i=1}^{n} Y_i$ decreases as n increases. Thus, $\frac{1}{n}\sum_{i=1}^{n} Y_i$ will concentrate around it's expected value which is equal to $\frac{1}{n} \cdot b\sum_{i=1}^{n} t_i$. This motivates that the MME of model 1 is

$$\hat{b} = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i}{\frac{1}{n}\sum_{i=1}^{n} t_i} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}. \tag{2.1}$$

Next, the maximum likelihood estimator will be investigated. However, first a bit of notation will be introduced. Note, that $\mathbf{Y} = (Y_1, \ldots, Y_n)$ which corresponds to the n random variables and $\mathbf{y} = (y_1, \ldots, y_n)$ which corresponds to the n observations of the experiment. Recall that

$$\begin{aligned}
\hat{b}_{MLE.1}(\mathbf{y}) &= \underset{b\in\mathbb{R}^+}{\operatorname{argmax}}\ \mathbb{P}_b(\mathbf{Y} = \mathbf{y}) \\
&= \underset{b\in\mathbb{R}^+}{\operatorname{argmax}} \prod_{i=1}^{n} (b \cdot t_i \mathbb{1}\{Y_i = 1\} + (1 - b \cdot t_i)\mathbb{1}\{Y_i = 0\}) \\
&= \underset{b\in\mathbb{R}^+}{\operatorname{argmax}}\ b^{\sum_{i=1}^{n} Y_i} \cdot \prod_{i:Y_i=1} t_i \cdot \prod_{i:Y_i=0} (1 - b \cdot t_i) \\
&= \underset{b\in\mathbb{R}^+}{\operatorname{argmax}} \sum_{i=1}^{n} Y_i \cdot \log b + \sum_{i:Y_i=1} \log t_i + \sum_{i:Y_i=0} \log(1 - b \cdot t_i).
\end{aligned}$$

To find the value of b which maximizes the argument above we investigate the stationary points. These are the points for which we have that $\frac{\partial}{\partial b}\mathbb{P}_b(\mathbf{Y} = \mathbf{y}) = 0$. This gives that

$$\frac{1}{b} \cdot \sum_{i=1}^{n} Y_i - \sum_{i:Y_i=0} \frac{t_i}{1 - b \cdot t_i} = 0. \tag{2.2}$$

Solving equation 2.2 for b is generally not analytically possible. However, note that we are interested in the case that $b \cdot t_i$ is very small. In that case equation 2.2 is approximately equal to

$$\frac{1}{b}\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} t_i$$

because if $b \cdot t_i$ is very small, the probability that $Y_i = 0$ is very high and $1 - b \cdot t_i$ is close to 1. This means that if $b \cdot t_i$ is very small the MLE is essentially the MME. However, the general MLE of the Bernoulli model is complicated. This is the motivation behind the Poisson model, which is discussed in the next section.

2.2 Estimator of b assuming model 2 (Poisson)

Model 1.2 states that the values of Y_i are assumed to come from a Poisson distribution with parameter $b \cdot t_i$. For small values of $b \cdot t_i$ this is an approximation of model 1.1 by the Poisson limit theorem. This theorem states that n trials of the Bernoulli distribution can be approximated with the Poisson distribution [2]. The Poisson distribution is also a discrete distribution which takes values within the positive integers. Thus, the value of Y_i could be larger than 1 in the case of the Poisson distribution. Therefore, the $\mathbb{P}(Y_i \geq 1)$ in the case of the Poisson model is compared to the $\mathbb{P}(Y_i = 1)$ in the case of the Bernoulli random variable. This can be compared in figure A.1

in the appendix.

The methods to determine the estimators in the case of model 1.1 will again be used. Observe that $\sum_{i=1}^{n} Y_i$ is again a Poisson distribution with mean $b\sum_{i=1}^{n} t_i$, since a sum of independent Poisson random variables is again a Poisson random variable. Therefore, the MME of this model is simply the same as in the Bernoulli case

$$\hat{b} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}.$$

Now, the maximum likelihood estimator will be investigated.

$$\begin{aligned}
\hat{b}_{\text{MLE.2}}(\mathbf{y}) &= \operatorname*{argmax}_{b\in\mathbb{R}^+} \mathbb{P}_b(\mathbf{Y}=\mathbf{y}) \\
&= \operatorname*{argmax}_{b\in\mathbb{R}^+} \prod_{i=1}^{n} e^{-b\cdot t_i}\mathbb{1}\{Y_i=0\} + b\cdot t_i e^{-b\cdot t_i}\mathbb{1}\{Y_i=1\} \\
&= \operatorname*{argmax}_{b\in\mathbb{R}^+} e^{-b\sum_{i=1}^{n} t_i}\cdot b^{\sum_{i=1}^{n} Y_i}\cdot \prod_{i:Y_i=1} t_i \\
&= \operatorname*{argmax}_{b\in\mathbb{R}^+} -b\sum_{i=1}^{n} t_i + \sum_{i=1}^{n} Y_i \log b + \sum_{i:Y_i=1} \log t_i.
\end{aligned}$$

Again to find the value of b which maximizes the argument above we investigate the stationary points. These are the points for which we have that $\frac{\partial}{\partial b}\mathbb{P}_b(\mathbf{Y}=\mathbf{y})=0$. This gives that

$$-\sum_{i=1}^{n} t_i + \sum_{i=1}^{n} Y_i \cdot \frac{1}{b} = 0.$$

This expression can be simplified for b which gives that the MLE of model 1.2 is

$$\hat{b} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}.$$

Thus, the estimator $\hat{b}$ will be considered as an estimator of b in the case that f is linear.

2.3 Quality of the estimator $\hat{b}$

The quality of an estimator is often expressed with help of the mean squared error (MSE). The MSE consists of the variance and the bias squared. Consequently, the MSE takes the variability of the data and the distance between the estimator and the true value of b into account. From [2] recall $\text{MSE}(\hat{b}) = \mathbb{E}_b[(\hat{b}-b)^2] = (\text{bias}_b(\hat{b}))^2 + \text{Var}_b(\hat{b})$. Note that the Y_i's are independent as defined in model 1.2.

The MSE of $\hat{b} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}$ will be determined. First of all, the bias of $\hat{b}$ is

$$\text{bias}_b(\hat{b}) = \mathbb{E}_b\left[\hat{b}\right] - b = \mathbb{E}_b\left[\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}\right] - b = \frac{\sum_{i=1}^{n} \mathbb{E}_b[Y_i]}{\sum_{i=1}^{n} t_i} - b = \frac{\sum_{i=1}^{n} b\cdot t_i}{\sum_{i=1}^{n} t_i} - b = b - b = 0.$$

Thus, the bias of $\hat{b}$ is 0, which means that the estimator is unbiased. The variance of $\hat{b}$ is

$$\mathrm{Var}_b\left(\hat{b}\right) = \mathrm{Var}_b\left(\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}\right) = \frac{1}{\left(\sum_{i=1}^{n} t_i\right)^2}\mathrm{Var}_b\left(\sum_{i=1}^{n} Y_i\right) = \frac{1}{\left(\sum_{i=1}^{n} t_i\right)^2}\sum_{i=1}^{n}\mathrm{Var}_b(Y_i).$$

Two cases will need to be distinguished. The first is the case if Y_i is a Bernoulli random variable. Then we find

$$\begin{aligned}\frac{1}{\left(\sum_{i=1}^{n} t_i\right)^2}\sum_{i=1}^{n}\mathrm{Var}_b(Y_i) &= \frac{1}{\left(\sum_{i=1}^{n} t_i\right)^2}\cdot\sum_{i=1}^{n}\left(bt_i\cdot(1-bt_i)\right)\\ &= \frac{b}{\sum_{i=1}^{n} t_i} - b^2\frac{\sum_{i=1}^{n}\left(t_i^2\right)}{\left(\sum_{i=1}^{n} t_i\right)^2}\\ &\leq \frac{b}{\sum_{i=1}^{n} t_i}.\end{aligned}$$

In the second case, the case that Y_i is a Poisson random variable, we find that

$$\begin{aligned}\frac{1}{\left(\sum_{i=1}^{n} t_i\right)^2}\sum_{i=1}^{n}\mathrm{Var}_b(Y_i) &= \frac{1}{\left(\sum_{i=1}^{n} t_i\right)^2}\cdot\sum_{i=1}^{n} bt_i\\ &= \frac{b}{\sum_{i=1}^{n} t_i}.\end{aligned}$$

Thus, it follows that in the case of model 1.1 $\mathrm{MSE}(\hat{b}) \leq b/\sum_{i=1}^{n} t_i$ and in case of model 1.2 $\mathrm{MSE}(\hat{b}) = b/\sum_{i=1}^{n} t_i$. Therefore, it can be concluded that the MSE is smaller or equal in the case where Y_i comes from the Bernoulli distribution.

Chapter 3

Bootstrap-method

The goal of this chapter will be exploring a method that will determine if the function f is linear. It will be assumed that $Y_i \sim Poisson(f(T_i))$, where the values of $T_i = t_i = \frac{i}{n}$ are given. In this chapter a test will be introduced that might be able to detect if the function f is linear.

3.1 Hypotheses Tests

To test is if f is linear, a hypothesis test will be used. To start a null hypothesis and an alternative hypothesis are needed. These are defined as follows:

$$H_0 : f(t) = b \cdot t \tag{3.1}$$
$$H_1 : \forall b > 0 \;\; \exists t > 0 \text{ such that } f(t) \neq b \cdot t \tag{3.2}$$

These hypotheses can also be formulated differently. Given model 1.2 one can reformulate the hypotheses in terms of $\lambda_i = f(t_i)$.

$$H_0 : \exists b > 0 \text{ such that } \forall i \in \{1, \ldots, n\} \; \lambda_i = b \cdot t_i$$
$$H_1 : \nexists b > 0 \text{ such that } \forall i \in \{1, \ldots, n\} \; \lambda_i = b \cdot t_i$$

A natural way to consider the hypotheses is the generalized likelihood ratio statistic. This test statistic looks at which of the hypothesis is more likely. Recall that the GLR statistic is defined as follows: If X has p.m.f. $f_\theta(\mathbf{x})$ with $\theta \in \Theta$ and $H_0 : \theta \in \Theta_0$, $H_1 : \theta \in \Theta_1$ then the generalized likelihood ratio statistic is defined as

$$\lambda(\mathbf{X}) = \frac{\sup_{\theta \in (\Theta_0 \cup \Theta_1)} f_\theta(\mathbf{x})}{\sup_{\theta \in \Theta_0} f_\theta(\mathbf{x})}.$$

See for instance book [2].

This problem is essentially a goodness-of-fit problem. Because we want to know if the proposed model agrees with the data. However, many goodness-of-fit tests assume that the data is independent and identically distributed. Therefore, it is complicated to apply such goodness-of-fit tests. One might consider approaches to still be able to use a traditional goodness-of-fit test, for example, averaging the data. However, the averaging of data might lead to the loss of information. This part is left for future research. In this report another approach will be investigated, namely a bootstrap-based method.

3.2 Bootstrap method

The bootstrap method is a simulation based approach. The main idea behind the bootstrap method is to generate surrogate data following the null hypothesis based on the observed data. This surrogate data is used to determine the behavior of the test statistic under the null hypothesis [6].

In other words new "data" is created using the null hypothesis and an estimation of the parameter from the original data. Then this new created surrogate data will be compared to the existing data. Finally, a conclusion about the null hypothesis will be made. This is done by comparing a test statistic of the original data and the same test statistic of the generated data (which is a surrogate of the test statistic under the null hypothesis).

The GLR will be used as explained above. The generalized likelihood ratio statistic is determined for the hypotheses 3.1 and 3.2, then we find

$$\begin{aligned}\lambda_{GLR}(\mathbf{Y}) &= \frac{\sup_{(\lambda_1,\ldots,\lambda_n)>0} \prod_{i=1}^{n} e^{-\lambda_i} \frac{\lambda_i^{Y_i}}{Y_i!}}{\sup_{b>0} \prod_{i=1}^{n} e^{-b\cdot t_i} \frac{(b\cdot t_i)^{Y_i}}{Y_i!}} \\ &= \frac{\prod_{i=1}^{n} e^{-Y_i} \frac{Y_i^{Y_i}}{Y_i!}}{\prod_{i=1}^{n} e^{-\hat{b}\cdot t_i} \frac{(\hat{b}\cdot t_i)^{Y_i}}{Y_i!}}. \end{aligned} \tag{3.3}$$

Note, that in the case that there are no restrictions on the parameters λ_i, every observation Y_i will have a corresponding parameter. The parameter corresponding to Y_i is λ_i. The MLE of λ_i is precisely Y_i because Y_i is a Poisson random variable. Furthermore, under the null hypothesis we simply evaluate the likelihood for b equal to the maximum likelihood estimator, as derived in section 2.2. This gave rise to the expression 3.3. The convention will be used that $0^0 = 1$.

For the bootstrap method we are interested in the comparison between the two test statistics. However, rather than working with equation 3.3 we will use a logarithmic transformation. This will help with numerical stability. The test statistic that is considered is given by

$$\begin{aligned}\Lambda(\mathbf{Y}) = 2\log(\lambda_{GLR}(\mathbf{Y})) &= 2\log\left(\frac{\prod_{i=1}^{n} e^{-Y_i} \frac{Y_i^{Y_i}}{Y_i!}}{\prod_{i=1}^{n} e^{-\hat{b}t_i} \frac{(\hat{b}t_i)^{Y_i}}{Y_i!}}\right) \\ &= 2\log\left(\prod_{i=1}^{n} e^{-Y_i+\hat{b}\cdot t_i} \left(\frac{Y_i}{\hat{b}\cdot t_i}\right)^{Y_i}\right) \\ &= \sum_{i=1}^{n} 2\left(-Y_i + \hat{b}\cdot t_i\right) + 2\log\left(\left(\frac{Y_i}{\hat{b}\cdot t_i}\right)^{Y_i}\right). \end{aligned} \tag{3.4}$$

The bootstrap approach with the generalized likelihood ratio statistic was implemented in R and the code will be included in the appendix A.2. A pseudo code is given in algorithm 1, to make the approach more clear.

3.3 Quality of bootstrap method

To determine the quality of this bootstrap method two aspects of this procedure will be investigated. Firstly, it will be investigated if the method is well calibrated (or at least conservative). This means that ideally the p-values should look like samples from a uniform distribution between 0 and 1 under the null hypothesis. Secondly, it will be investigated how powerful this method is. Namely, how often is the null hypothesis rejected when $f(t)$ is not linear for all t.

Algorithm 1 Bootstrap method

Input: $(y_i, t_i)_{i=1}^n$ and K: number of bootstrap repetitions
Output: p-value

1: Compute $\hat{b}^{(0)} = \sum_{i=1}^{n} y_i / \sum_{i=1}^{n} t_i$
2: Compute the test statistic as in 3.4 $\Lambda^{(0)} = \sum_{i=1}^{n} 2\left(-y_i + \hat{b}^{(0)} \cdot t_i\right) + 2\log\left(\left(\frac{y_i}{\hat{b}^{(0)} \cdot t_i}\right)^{y_i}\right)$
3: **for** k from 1 until K **do**
4: Create $y_i^{(k)}$ = vector of n independent generated samples from the Poisson distribution with parameter $\hat{b}^{(0)} \cdot t_i$
5: Compute $\hat{b}^{(k)} = \sum_{i=1}^{n} y_i^{(k)} / \sum_{i=1}^{n} t_i$
6: Compute $\Lambda^{(k)} = \sum_{i=1}^{n} 2\left(-y_i^{(k)} + \hat{b}^{(k)} \cdot t_i\right) + 2\log\left(\left(\frac{y_i^{(k)}}{\hat{b}^{(k)} \cdot t_i}\right)^{y_i^{(k)}}\right)$
7: **end for**
8: p-value $= \frac{1}{K} \cdot \sum_{k=1}^{K} \mathbb{1}\{\Lambda^{(k)} > \Lambda^{(0)}\}$

b	n	proportion p-values≤ 0.05
10	100	0.0512
1	100	0.0400
0.5	100	0.0146
10	1000	0.0494
1	1000	0.0408
0.5	1000	0.0116

Table 3.1: Proportion p-values≤ 0.05

3.3.1 Quality of bootstrap under null hypothesis

It was investigated if the bootstrap method is well calibrated. This means that under the null hypothesis the p-values come from a uniform distribution between (0,1) (proposition 4.1 in [2]) and that with level of rejection α the proportion of rejected trials (type 1 errors) should be equal to α. If this is not the case one would like the method to be conservative (reject less than a proportion of α). Thus, one would like to conclude that
$\forall \alpha > 0$ we have that $\mathbb{P}_0(\text{p-value} \leq \alpha) \leq \alpha$.

The bootstrap method is tested in the software R. Data was created under the null hypothesis where $t_i = \frac{i}{n}$. Different values of b and n are implemented. Then the bootstrap method was applied with $K = 1000$ bootstrap samples. This is repeated 5000 times. The results can be observed in Table 3.1 and Figure 3.1.

In Figure 3.1 we plot the sorted p-values for the 5000 repetitions and each experimental condition. In the appendix the histograms can also be observed in Figure A.2. However, the plots in Figure 3.1 give a more precise picture since the number of bins in a histogram can change the picture quite much. If the testing procedure is well calibrated these plots should look like a straight line ranging from 0 to 1 (in red). Observe that for larger values of b the method seems to be better calibrated since the sorted p-values seem to follow the uniform distribution between (0,1). However, for smaller values of b the p-values do not seem uniform anymore. This can be explained by the fact that the bootstrap method relies on a good estimate of the parameter b. Since it constructs surrogate data using the null hypothesis and the estimate, the method does not work very well if the estimate is not very good. This is the case when b is smaller, because the data carries less information. It carries less information because the probability that your observation

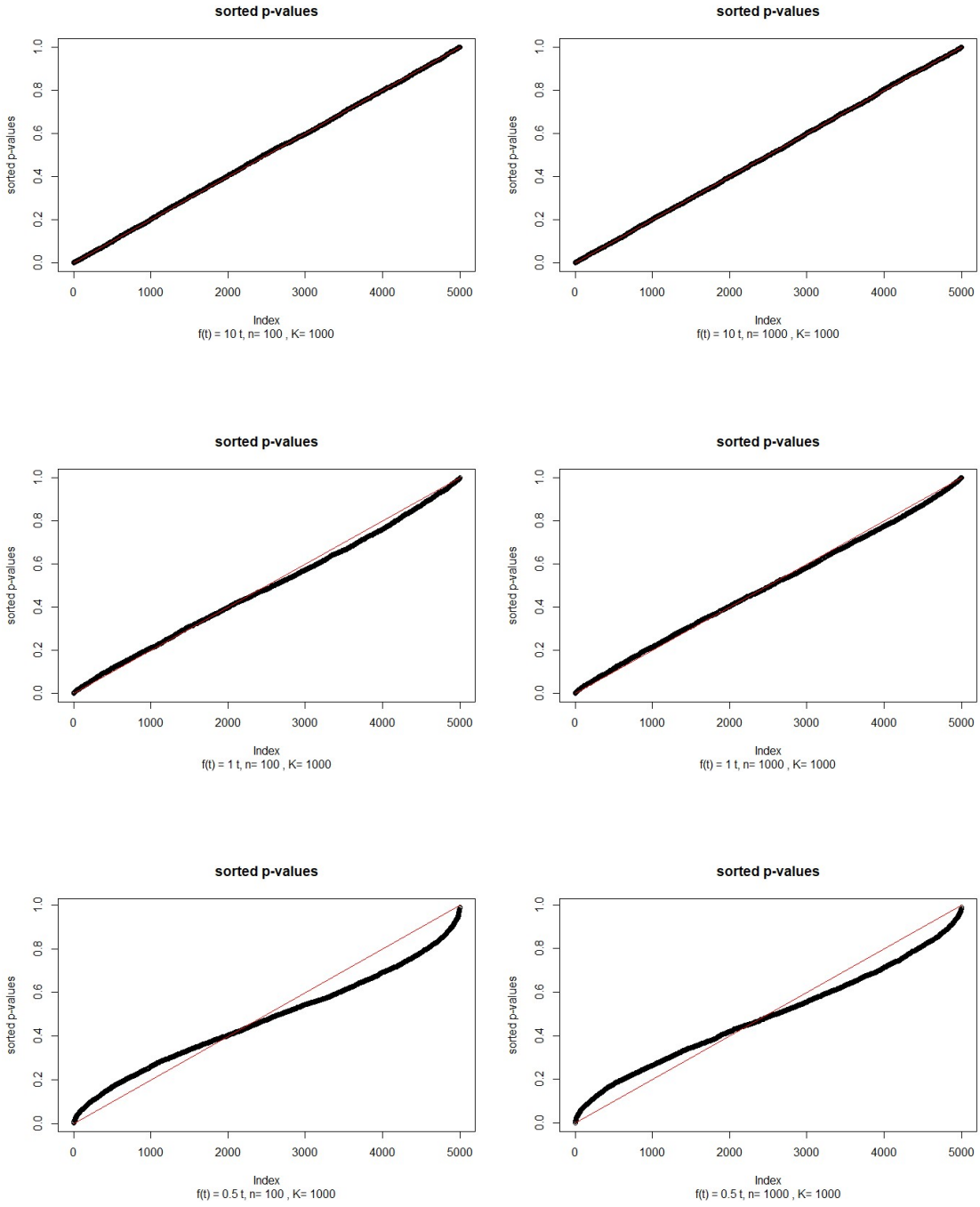


Figure 3.1: Plots of p-values for multiple values of b and n

is equal to 0 is very large. This results into data with a lot of zeros and just a few ones. This results into an estimate of b which is not as good as in the case where b is larger. One can observe that for a larger value of n and thus more data the distance from the red line gets smaller (less far from uniform). This is as expected, because if there is more data the estimate of b will most likely be better. The method seems to be conservative. The p-values namely increase faster than expected (uniform) this means that there are fewer p-values which are less than 0.05. This implies that there were rejected fewer cases than the expected proportion of 0.05 which means that for these values of b, n and K the method is conservative. This is on the one hand good, since the null hypothesis is not rejected too fast. However, when talking about goodness-of-fit testing this might also be undesirable. If a goodness-of-fit test is conservative it will be harder to reject the null hypothesis. When using a goodness-of-fit test one is often looking for evidence against the null model. If the method is conservative the result of the test might misleadingly lead to the use of this linear model. The power of this test will be investigated in the next section.

In Table 3.1 one can observe the proportion of repetitions that were rejected with $\alpha = 0.05$. Important to note is that these are the results after 5000 repetitions. If the experiment would be repeated, other (but similar) results might arise. The proportions are close to 0.05 or lower. This would indicate that the method if is well calibrated (due to enough information from the data and enough data) then the proportion of rejected repetitions is close to 0.05. However, if the method does not have a lot of information as in the cases of $b = 1$ or $b = 0.5$ then the method seems to be conservative.

Overall, it can be concluded that the bootstrap method is a well-calibrated test when investigating it under the null hypothesis under the constraint that there is enough information from the data. This could form a problem in the case of the quantum technology since the value of b will probably be very small. This implies that a lot of data is needed to be able to apply this method.

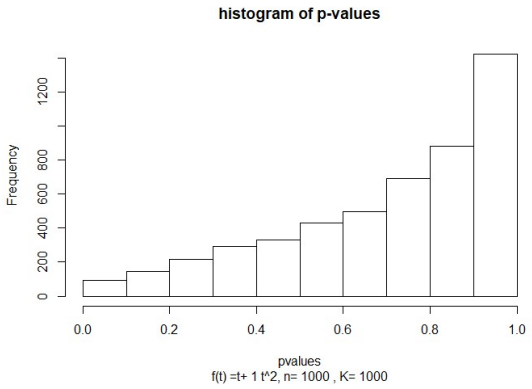
3.3.2 Quality of bootstrap under alternative hypothesis

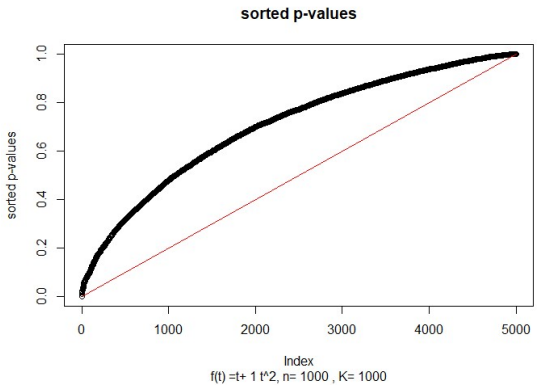
In this section the power of the test will be investigated. This means that data coming from the alternative hypothesis will be tested with the bootstrap method. One would like that if data does not come from the null hypothesis the method would reject H_0. However, it is very unlikely that a test would always reject data that comes from an alternative hypothesis. Therefore, the interest will lie with the power of the method, this is the proportion of p-values< 0.05. Since the alternative hypothesis is not one specific function f not all possible alternative hypothesis can be tested. Therefore, three cases are investigated:

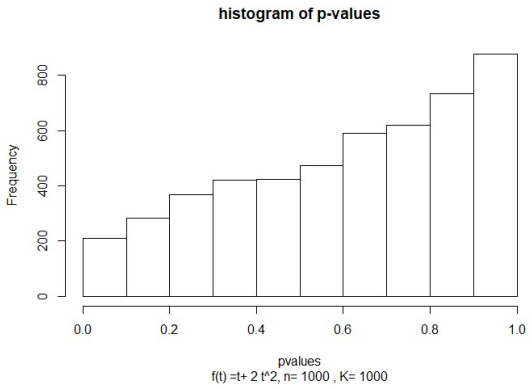
1. $f(t) = b \cdot t + c \cdot t^2$ where c is a constant
2. $f(t) = \begin{cases} b_0 t & \text{if } t < t_0 \\ b_1(t - t_0) + b_0 t_0 & \text{if } t \geq t_0 \end{cases}$.
3. Y_i is the absolute value of a sample from the normal distribution with mean $b \cdot t_i$ and standard deviation 1. Note that the method (in some cases) does not work if Y_i is negative, since the parameter of a Poisson distribution has to be positive. Therefore, the absolute value is considered.

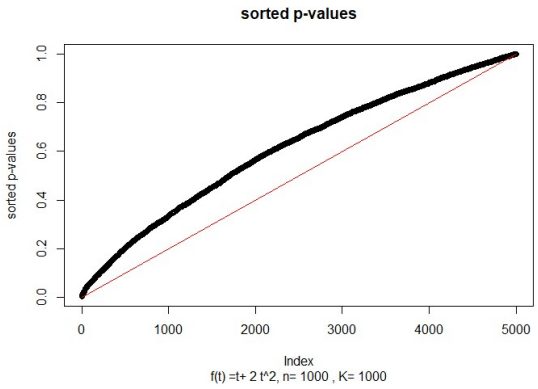
Data will be created following the alternative hypothesis of one of these cases. In case 3 the data does not come from the Poisson distribution and the data is not discrete. However, the data can be implemented as input for the bootstrap method. This alternative hypothesis is considered to investigate what happens if wrong input is used or if the data is not from the Poisson distribution.

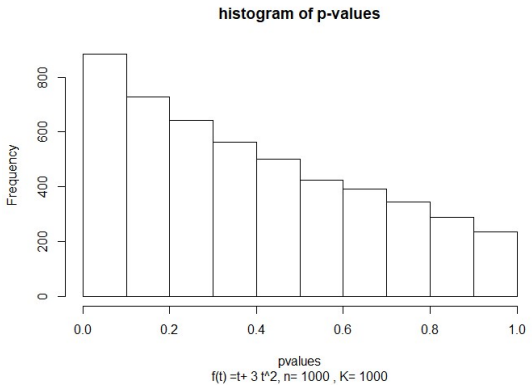
First case 1 will be investigated. In this case the function f is not linear as the null hypothesis states but it contains a quadratic term. This alternative hypothesis was implemented in R with $K = 1000$, $n = 1000$, $b = 1$ and different values of c these experiments were repeated 5000 times.

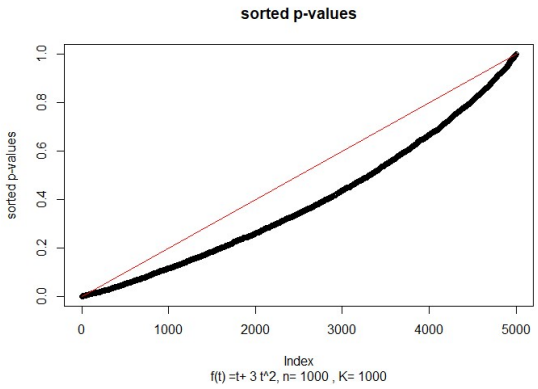

histogram of p-values
Frequency
0
200
400
600
800
1200
0.0
0.2
0.4
0.6
0.8
1.0
pvalues
f(t) =t+ 1 t^2, n= 1000 , K= 1000


sorted p-values
sorted p-values
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
5000
Index
f(t) =t+ 1 t^2, n= 1000 , K= 1000


histogram of p-values
Frequency
0
200
400
600
800
0.0
0.2
0.4
0.6
0.8
1.0
pvalues
f(t) =t+ 2 t^2, n= 1000 , K= 1000


sorted p-values
sorted p-values
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
5000
Index
f(t) =t+ 2 t^2, n= 1000 , K= 1000


histogram of p-values
Frequency
0
200
400
600
800
0.0
0.2
0.4
0.6
0.8
1.0
pvalues
f(t) =t+ 3 t^2, n= 1000 , K= 1000


sorted p-values
sorted p-values
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
5000
Index
f(t) =t+ 3 t^2, n= 1000 , K= 1000

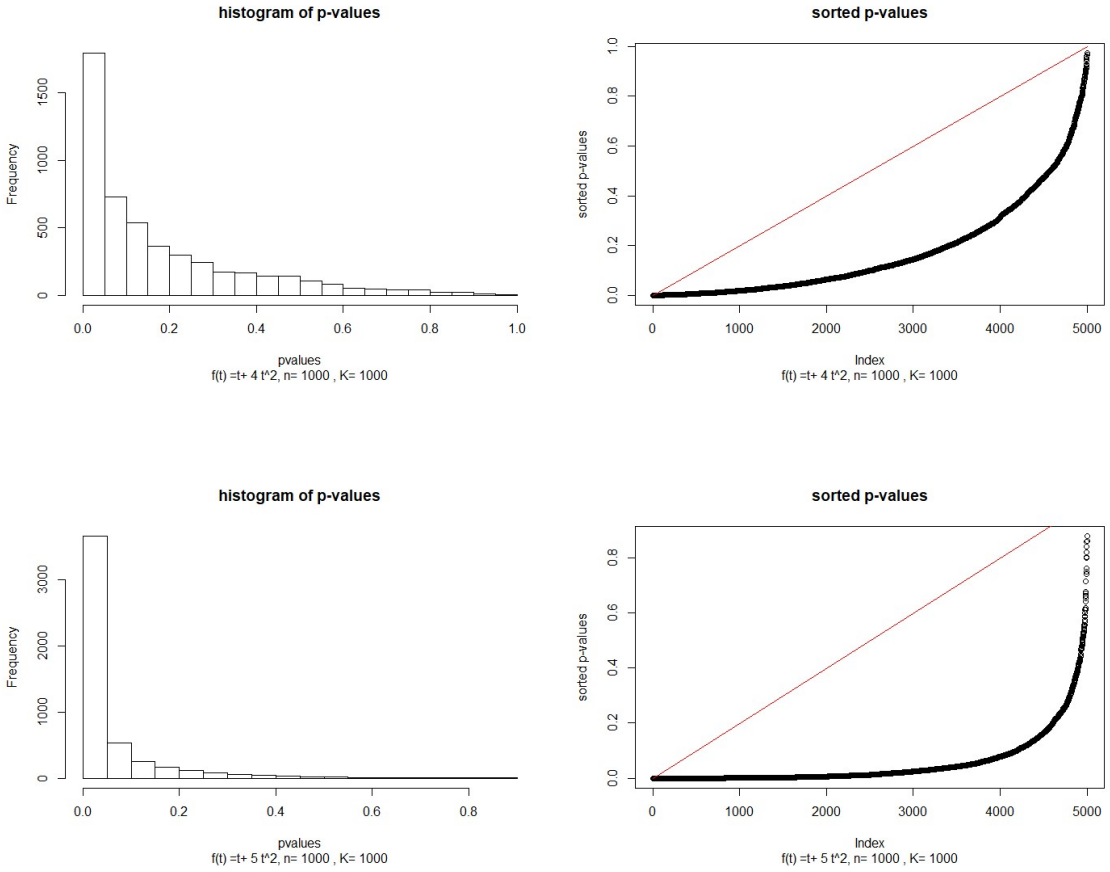


Figure 3.2: Plots of p-values for data with $f(t) = t + ct^2$

c	proportion p-values$\leq$ 0.05
1	0.0048
2	0.0146
3	0.0970
4	0.3558
5	0.7290

Table 3.2: Proportion p-values$\leq$ 0.05 for $f(t) = t + ct^2$

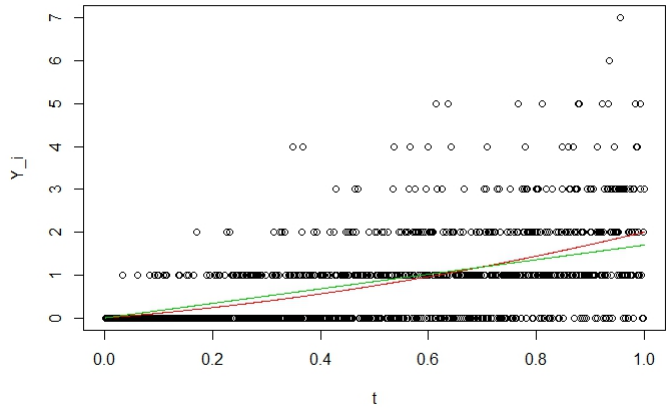


Figure 3.3: Comparison $f(t)$ and $\hat{b} \cdot t$
red: $f(t)$ and green: $\hat{b} \cdot t$

b_0	b_1	proportion p-values≤ 0.05
1	2	0.006
10	20	0.8928

Table 3.3: Proportion p-values≤ 0.05 for $f(t)$ with a jump

The results can be observed in Figure 3.2 and Table 3.2. In the figures in 3.2 one can observe the p-values for different functions $f(t) = t + c \cdot t^2$. In Table 3.2 the proportion of rejections of all repetitions with $\alpha = 0.05$ can be observed. For the alternative hypothesis it is desirable that this value is as high as possible. One can observe that for $c = \{3, 4, 5\}$ the values of the p-values are relatively smaller than in the case of a uniform distribution and that the proportion of rejections is higher than 0.05. However, for $c \in \{1, 2\}$ the results are quite counter intuitive since the proportion of rejection is very small (even smaller than in the case of the null hypothesis). This would imply that the case when $f(t) = t$ will be rejected more often than the case when $f(t) = t + t^2$. This, is a very curious observation. It can partially be explained by the fact that the test statistic for the null hypothesis is estimated with the bootstrap. In Figure 3.3 the data points, $f(t) = t + t^2$ and the estimated $\hat{b} \cdot t$ will be given. Here it becomes clear that the two functions are very close and that it can indeed be very hard to decide which of the two is more likely. However, more research should be done into this topic if this method wants to be explored further.

$$\text{In case 2 } f(t) = \begin{cases} b_0 t & \text{if } t < t_0 \\ b_1(t - t_0) + b_0 t_0 & \text{if } t \geq t_0 \end{cases}.$$

This function has one derivative (b_0) until some point which is $t = t_0$ and after that another derivative (b_1). This function is thus a continuous function with a discontinuous derivative. A function like this with $t_0 = 0.5, b_0 = 10$ and $b_1 = 20$ can be observed in Figure 3.4. For $(b_0, b_1) = (1, 2)$ and $(b_0, b_1) = (10, 20)$ the function was tested with the bootstrap method. The results can be observed in Figure 3.5 and Table 3.3.

The same behavior as for the case of f having a quadratic term can be observed. For smaller values of the parameters (and therefore $f(t)$) the method has the unexpected behavior of only rejecting a very small amount of repetitions. In the case of higher parameters the power of the method seems to be much better. This can be explained by the fact that if the parameters are higher the data does contain more information (more non-zero data points).

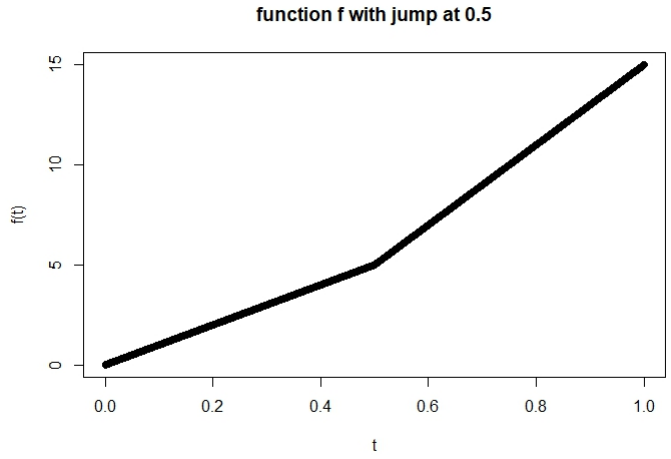


Figure 3.4: $f(t) = 10 \cdot t$ for $t \leq 0.5$ and $f(t) = 20 \cdot (t - 0.5) + 10 \cdot 0.5$ for $t > 0.5$

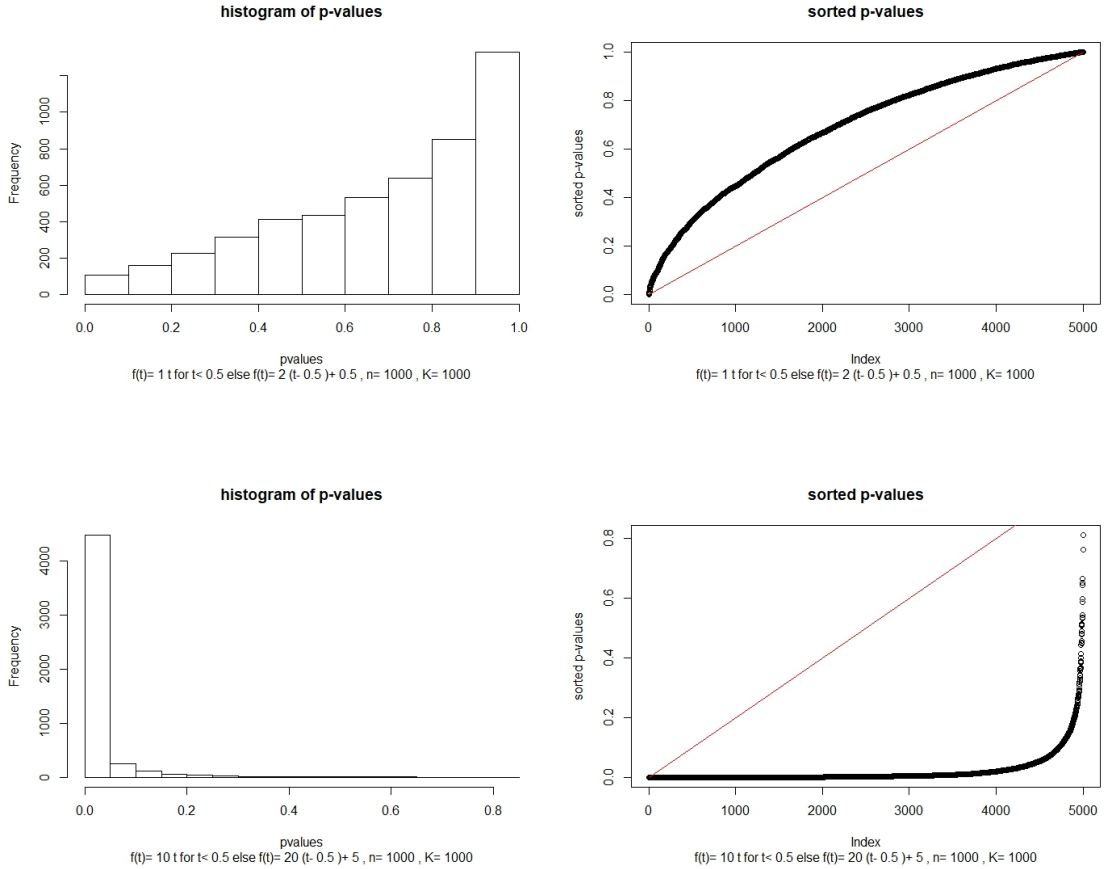


Figure 3.5: Plots of p-values for data from normal distribution

In case 3 we have that Y_i is the absolute value of a normal sample with mean $b \cdot t_i$ and standard deviation 1.

To try what would happen if the data does not come from a Poisson distribution the absolute value of generated data from the normal distribution was used as input to the bootstrap method. The normal distribution has mean $b \cdot t$ and variance 1. This means that the data looks like in Figure 3.6, the red line is $b \cdot t$.

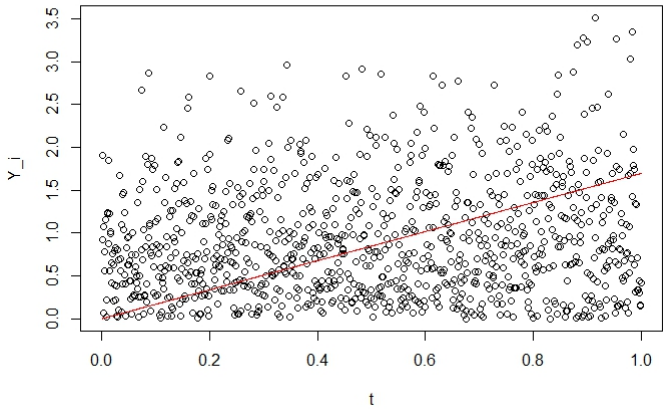


Figure 3.6: Absolute value of normal data with mean $1 \cdot t_i$ and standard deviation 1

Note that the null hypothesis is that the function $f(t) = b \cdot t$ and that it is assumed that the data comes from the Poisson distribution. The results from the bootstrap method when this normal data is applied can be found in Figure 3.7.

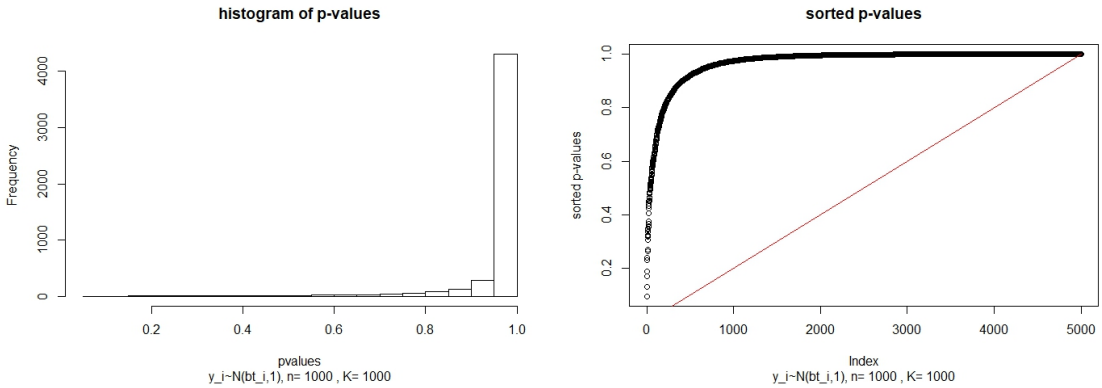


Figure 3.7: Plots of p-values for data from normal distribution

In Figure 3.7 one can see that the p-values increase very fast and that they are much higher than expected when data comes from an alternative hypothesis. This seems quite counter intuitive. However, it can be explained by the two following facts. Recall that the method tests if $\exists b > 0$ such that $\forall i \in \{1, ..., n\}$ $\lambda_i = b \cdot t_i$. Observe that under the null hypothesis (and assuming

the Poisson distribution) $\mathbb{E}[Y_i] = f(t_i) = b \cdot t_i$. The alternative hypothesis has the same expected value. Furthermore, for the null hypothesis the variance is equal to $b \cdot t_i$ and for the alternative hypothesis the variance is equal to 1. This implies that the data points are centered around the function $f(t) = b \cdot t$. This results in high p-values.

Overall, the power of the test is dependent on the underlying model of the data and on how much information there is in the data. More research should be done into why these p-values are so high in some cases, and how to prevent these high p-values in case of some alternative hypotheses.

To test if f is linear a bootstrap method was applied. This method worked well in the cases of a large value of b and if n was very large. Furthermore, the method seemed to have a good power in case of a large deviation from the linear case. But, if the function was not far from linear, the method has some problems. Since we are interested in the case that b is very small and the power of the test depends on the deviation from a linear f, this method is not recommended in the way it is stated in this report. Some adjustments might improve this method. This is left for future research.

Chapter 4

Intervals-estimator

In this chapter we move away from the assumption that $f(t)$ is linear. Instead we will assume that $f(t) = b \cdot t + o(t)$ as $t \to 0$. This implies that for small values of t the function is close to linear, but otherwise the function can be arbitrary. The idea of how to find the estimate came from the behavior of the estimator derived in Chapter 2. Recall that the estimator of b, as derived in Chapter 2, is $\hat{b} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}$.

4.1 Background

In this section the background and basic notation of the method to estimate b will be given. Again it is assumed that $T_i = t_i = \frac{i}{n}$ is given. In this case the values of t_i are ordered. However, if one would chose to select t_i in another way, it is important to let $t_1 \leq t_2 \leq \cdots \leq t_n$, such that the first k datapoints correspond to the smallest k values of $(t_i)_{i=1}^n$.

4.1.1 Behavior of $\hat{b}_k$

This method arose when thinking about the following question: what happens if you do not take all data for the estimator but only a part of it? In this chapter it will be investigated what happens to the estimator $\hat{b}$ if only the data until some point $k \in \{1, ..., n\}$ will be used for the estimator. The reasoning behind this idea is that the assumption is that f is only linear for small values of t. This would imply that only the data for small values of t should be used for the estimate.

Let us define

$$\hat{b}_k = \frac{\sum_{n=1}^{k} Y_i}{\sum_{n=1}^{k} t_i} \text{ where } k \in \{1, ..., n\}.$$

In Figure 4.1 a sequence of $\hat{b}_k$ where $k \in \{1, ..., 10000\}$ can be observed. Note that Y_i follows a Poisson distribution with parameter $f(t)$ which is linear under the null hypothesis. The values of t_i are defined as follows $t_i = \frac{i}{n}$, thus it is a partition of the interval $[0, 1]$. In Figure 4.1 it is assumed that $Y_i \sim \text{Poisson}(t_i)$ or $Y_i \sim \text{Poisson}(t_i + t_i^2)$. In Figure 4.1 it can be observed that in the case of the linear function f the $\hat{b}_k$'s seem to converge to the right value 1. For small values of k the $\hat{b}_k$ seem to oscillate quite much. However, for $k > 2000$ the oscillations seem to be much smaller and $\hat{b}_k$ is close to 1. In the case of $f(t) = t + t^2$ the value converges to approximately 1.6. For small values of k, approximately for $k < 2000$ one can observe that the values of $\hat{b}_k$ oscillate quite much. For larger values of k this is no longer the case. In the case of the f having a quadratic

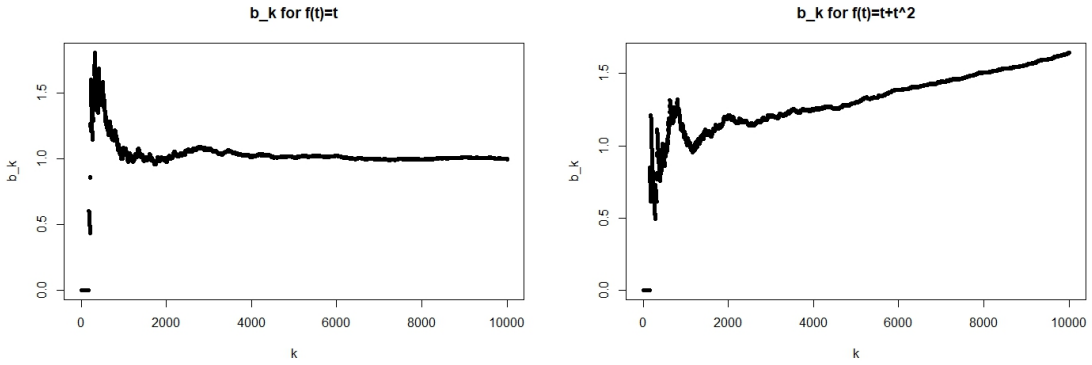


Figure 4.1: $\hat{b}_k$ for Y_i under the null and the alternative

term, for k2000 the $\hat{b}_k$ already seems much closer to 1. This gave rise to the suspicion that the behavior of $\hat{b}_k$ contains information up to which point the data should be considered for the estimator.

4.1.2 Confidence interval with normal approximation

The behavior of $\hat{b}_k$ contains some information about the value of b and until where the function might be linear. Therefore, we would like to know how certain we are about the values of $\hat{b}_k$. For example, for small values of k there is not much data. Therefore, the value of $\hat{b}_k$ might not be very good. To better understand this we proceed by deriving confidence intervals for b based on the first k datapoints, under the assumptions that $f(t)$ is linear up to that point. The distribution of the estimator $\hat{b}_k = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}$ can be approximated with the normal distribution due to the central limit theorem and the fact that Y_i is a Poisson random variable. Note that under the assumption that f is linear we have

$$\mathbb{E}\left[\hat{b}_k\right] = b, \tag{4.1}$$

and

$$\mathrm{Var}\left(\hat{b}_k\right) = \mathbb{E}\left[\left(\frac{\sum_{i=1}^{k} Y_i}{\sum_{i=1}^{k} t_i} - b\right)^2\right] = \frac{b}{\sum_{i=1}^{k} t_i}.$$

These calculations follow the same steps as in section 2.3. The central limit theorem and the fact that Y_i is a Poisson random variable implies that $\hat{b}_k$ is approximately $\mathcal{N}\left(b, \frac{b}{\sum_{i=1}^{k} t_i}\right)$, under the assumption that n is large and that f is linear.

This implies that $\mathbb{P}\left(b \in \left[\hat{b}_k - z_{\alpha/2}\sqrt{\frac{\hat{b}}{\sum_{i=1}^{k} t_i}}, \hat{b}_k + z_{\alpha/2}\sqrt{\frac{\hat{b}}{\sum_{i=1}^{k} t_i}}\right]\right) \approx 1 - \alpha.$

To get an estimate of b one would like all confidence intervals to be simultaneously valid. Let us define the confidence intervals of b using the datapoints where $i \leq k$ as $[L_k, U_k]$ where

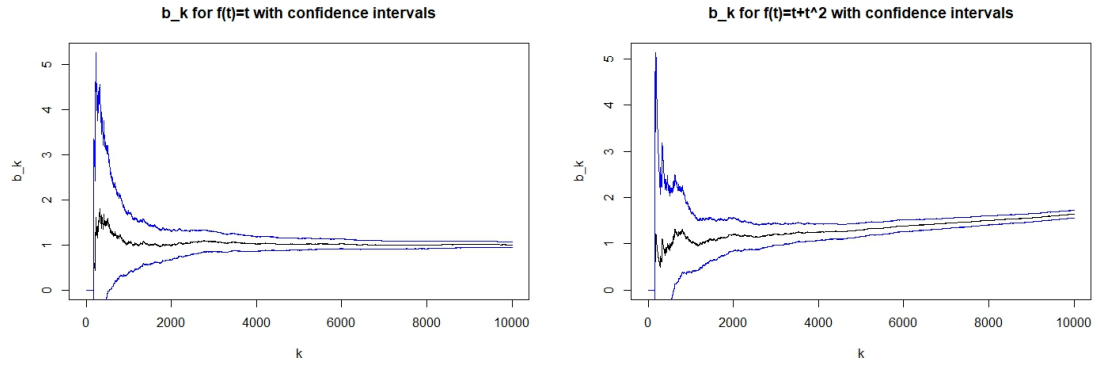


Figure 4.2: b_k with confidence intervals

$L_k = \hat{b}_k - z_{\alpha/2}\sqrt{\frac{\hat{b}_k}{\sum_{i=1}^{k} t_i}}$ and $U_k = \hat{b}_k + z_{\alpha/2}\sqrt{\frac{\hat{b}_k}{\sum_{i=1}^{k} t_i}}$. One would like to conclude that

$$\mathbb{P}\left(\forall k : b \in [L_k, U_k]\right) \geq 1 - \alpha.$$

This can be concluded under the assumption that f is linear and if α is chosen in a particular way .

Let $[L_k, U_k]$ be the confidence interval of b. Then

$$\begin{aligned}
1 - \mathbb{P}\left(\forall k : b \in [L_k, U_k]\right) &= \mathbb{P}\left(\exists k : b \notin [L_k, U_k]\right) \\
&= \mathbb{P}\left(\bigcup_k \{b \notin [L_k, U_k]\}\right) \\
&\leq \sum_{k=1}^{n} \mathbb{P}\left(b \notin [L_k, U_k]\right) \quad \text{(by Boole's law)} \\
&\text{now let } [L_k, U_k] \text{ be the } \frac{\alpha}{n} \text{ confidence interval of } b \text{ using } (y_i, t_i)_{i=1}^{k}, \text{ then} \\
&= \sum_{k=1}^{n} \mathbb{P}\left(b \notin \left[\hat{b}_k - z_{\alpha/(2n)}\sqrt{\hat{b}_k / \sum_{i=1}^{k} t_i}, \hat{b}_k + z_{\alpha/(2n)}\sqrt{\hat{b}_k / \sum_{i=1}^{k} t_i}\right]\right) \\
&= \sum_{k=1}^{n} \frac{\alpha}{n} = \alpha.
\end{aligned}$$

This implies that if, for example, $\alpha = 0.05$ and $[L_k, U_k]$ is a $\frac{\alpha}{n}$ confidence intervals of b using the first k data points then $\mathbb{P}\left(\forall k : b \in [L_k, U_k]\right) \geq 0.95$ under the assumption that f is linear. Note that α needs to be chosen.

In Figure 4.2 the exact same b_k as in Figure 4.1 can be observed. Here with the confidence intervals $[L_k, U_k]$. Note that the it is chosen that $\alpha = 0.05$. This implies that $[L_k, U_k]$ are the $\frac{0.05}{n}$ confidence intervals of b. Here again $n = 10000$. It can be observed that the confidence intervals in Figure 4.2 get smaller for larger values of k. This can be explained by the fact that more data is used and therefore $\sum_{i=1}^{k} t_i$ increases.

In the case of $f(t) = t + t^2$ if one would plot a line at $b = 1$ one would be able to see that for large values of k, $\hat{b}_k$ does not fit in all confidence intervals anymore. This is due to the fact that the f is not linear, which leads to inconsistent confidence intervals.

The method that will determine an estimate of b will use two observations. The first is that as in Figure 4.2 the confidence intervals of the estimator are large in the beginning and get smaller for larger k. Moreover, for smaller k the confidence intervals should be more reliable, as they are build upon the assumption that f is linear. How to decide for which k the estimation of b is close to the true value of b, will be described in the following section.

4.2 Intervals method

The ideas behind the method were explained in section 4.1. One would like to select the $\hat{b}_k$ that is closest to the value of b. On the one hand one would like to select a k which is as large as possible. If k is large a lot a data is used and the confidence interval is small. On the other hand one would like to select a k which is fairly small since for small values of k the function $f(t)$ is more likely to be linear and thus the data is reliable. Therefore, a compromise will be made. Note that all intervals are simultaneously valid.

Therefore, the estimator that the intervals method will give is $\hat{b}_{\text{intervals}} = \hat{b}_{\hat{k}}$ where $\hat{k}$ will be selected in the following way. A set S will be defined as the set of all $k \in \{1, ..., n\}$ such that $\hat{b}_k$ is in all previous confidence intervals. Note that $[L_k, U_k]$ is the $\frac{\alpha}{n}$ confidence interval of b based on the first k datapoints. So,

$$\begin{aligned} S &= \{k \in \{1, ..., n\} : \forall i \leq k \; L_i \leq \hat{b}_k \; \wedge \; U_i \geq \hat{b}_k\} \\ &= \{k \in \{1, ..., n\} : \; \max_{i \leq k}(L_i) \leq \hat{b}_k \quad \wedge \; \min_{i \leq k}(U_i) \geq \hat{b}_k\}. \end{aligned}$$

Then $\hat{k}$ will be defined as the maximum of this set S. Thus

$$\hat{k} = \max\{S\} = \max_{k \in \{1,...,n\}} \{k : \max_{i \leq k}(L_i) \leq \hat{b}_k \leq \min_{i \leq k}(U_i)\}.$$

Finally, the intervals estimator is defined as $\hat{b}_{\hat{k}}$ where $\hat{k} = \max_{k \in \{1,...,n\}} \{k : \max_{i \leq k}(L_i) \leq \hat{b}_k \leq \min_{i \leq k}(U_i)\}$. Note that $\hat{k}$ is as large as possible while $\hat{b}_{\hat{k}}$ is greater than all previous lower bounds and smaller than all previous upper bounds.

This method is implemented using cumulative functions. In the simulations it will be used that if $\sum_{i=1}^{j} Y_i < 5$ for some j then $U_j = 1000$ and $L_j = 0$. This will make sure that the first few confidence intervals do not effect the outcome. This is done as the central limit approximation used for the construction of the confidence intervals is only valid when $\sum_{i=1}^{j} Y_i$ is large enough. In this way we are able to ensure that the interval, where this approximation is not yet valid, has enough coverage.

To implement this method two parameters need to be chosen. The first is α which will determine the confidence intervals. The second is the value of $\sum_{i=1}^{k} Y_i$ before which the confidence intervals will be set to $[0, 1000]$.

4.3 Quality of intervals method

The quality and behavior of the intervals method will also be investigated with some simulations. All simulations were done in the software R, the code will be included in the appendix A.3. Three cases will be investigated. The first is what happens if the function f is linear. Then what happens if f is not linear, the cases discussed will be f with a quadratic term and f with a jump in derivative. These cases were also discussed in Chapter 3.3.2. The quantities that will be of interest here are $\hat{b}_{intervals}$, $\hat{k}$ and the MSE. Note that model 1.2 is assumed and T_i is equal to $t_i = \frac{i}{n}$.

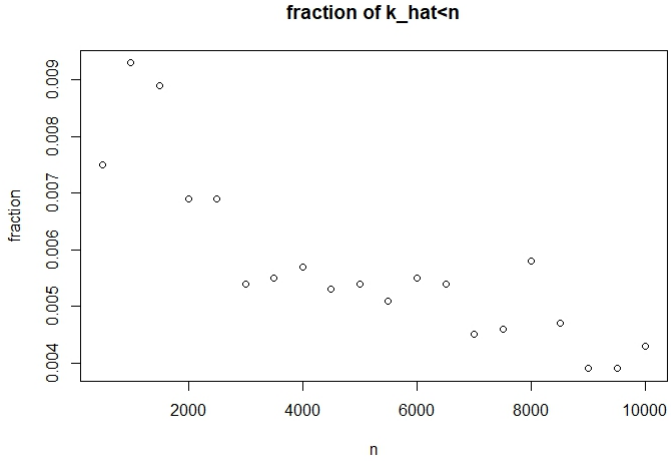


Figure 4.3: Proportion of $\hat{k} < n$

Case 1 f linear
If the function f is linear the estimator is unbiased and the confidence intervals around $\hat{b}_k$ get smaller for larger values of k. In Figure 4.2 one can observe that for the case of a linear f the smaller confidence intervals are contained in the larger ones. This implies that it is expected that b_n fits in all confidence intervals. Therefore, $\hat{k}$ will most likely be equal to n. Then the intervals estimator is the same as the estimator found in Chapter 2. This is preferable, since all data is used and this improves the estimator in the case of a linear function f.

However, what happens if $\hat{k} < n$ (which is the same as $\hat{k} \neq n$) and how often does this happen? The method was executed for different values of n the results can be observed in Figure 4.3. n ranges from 0 to 10000 and these experiments were repeated 10000 times with $b = 1$ and the confidence intervals are constructed with $\alpha = \frac{0.05}{n}$. For every n the fraction of repetitions where $\hat{k} < n$ is determined. It can be observed in Figure 4.3 that the fraction of repetitions where $\hat{k} \neq n$ is very small and gets smaller if n grows. This can be explained by the fact that if there is more data the value of $\hat{b}_k$ will be more likely to be close to b. Recall that $\mathbb{P}(\forall k : b \in [L_k, U_k]) \geq 1 - \alpha$ which means that if $\hat{b}_k$ is closer to b the probability that it is in all previous intervals is high.

The next question that might be asked is what $\hat{k}$ is if it is not equal to n. A histogram is be given in Figure 4.4 with the values of $\hat{k} < n$ for $n = 1000$, $b = 1$ and 1 million repetitions.

The fraction of experiments where $\hat{k} < 1000$ was 0.007928. It can be observed that for values close to $n = 1000$ there are respectively more $\hat{k}$'s selected. This can be explained by the fact that $\hat{b}_k$ is very likely to be very close to b for large k. Furthermore, the confidence intervals get much smaller which might lead to the last few $\hat{b}_k$ not being in all previous confidence intervals. Moreover, between 0 and 300 less $\hat{k}$'s are selected. This can be explained by the fact that for small values of k the confidence intervals are very large. This implies that the probability that b is contained in such a confidence interval is fairly high. Thus, the probability that there exists a $\hat{b}_k$ such that it is not contained in these intervals is very small.

Recall that $\hat{b}_k$ is unbiased for all k under the assumption that f is linear, as derived in equation 4.1. This implies that the MSE under the assumption that f is linear is equal to the variance of $\hat{b}_{\hat{k}}$, which depends on the choice of $\hat{k}$.

An approximation of the MSE was determined.

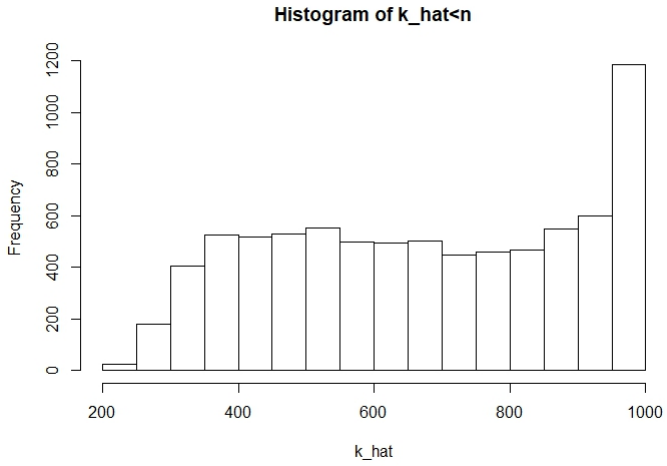


Figure 4.4: Distribution of $\hat{k}$ for $\hat{k} \neq n$

f	n	mean $\hat{b}$	mean $\hat{b}_{intervals}$	mean $\hat{k} = n$	$\text{MSE}_{\hat{b}}$	$\text{MSE}_{\hat{b}_{intervals}}$
$t + t^2$	1000	1.6667	1.4913	0.1268	0.4479	0.2676
$t + t^2$	10000	1.6668	1.3224	0	0.4449	0.1116
$t + 5t^2$	1000	4.3330	2.2081	0	11.1179	1.5886
$t + 5t^2$	10000	4.3335	1.7327	0	11.1128	0.5761

Table 4.1: Results from intervals method for f with a quadratic term

The average over all repetitions of $(\hat{b}_{intervals} - b)^2$ was determined and of $(\hat{b} - b)^2$. In this case $n = 1000$, $b = 1$ and there were 100,000 repetitions. For $\hat{b}_{intervals}$ the approximation of the MSE is 0.002072 and for $\hat{b}$ it is 0.002007. This means that they are very close but the approximation of $\hat{b}$ is a bit smaller.

Case 2 f with a quadratic term
The value of $\hat{k}$ will be around the point where the confidence intervals become inconsistent. In Figure 4.2 one can observe that the confidence intervals become inconsistent after a while, meaning the later confidence intervals are not contained in the earlier ones. If the confidence intervals are inconsistent the method will stop somewhere around there since a $\hat{b}_k$ in an inconsistent interval will very likely not be contained in all previous intervals. This behavior can be observed in Figure 4.5. Note that these confidence intervals were created with $\alpha = \frac{0.05}{n}$ and with $n = 1000$ and $n = 10,000$.
In Figure 4.5, one can observe one run of the intervals method. The determined $\hat{b}_k$ and $[L_k, U_k]$ are plotted. The red lines indicate the $\hat{k}$ and $\hat{b}_{\hat{k}}$ that the intervals method found. For these runs it can be observed that $\hat{k} < n$ and $\hat{b}_{intervals} < \hat{b}$.

Next, it will be investigated if this is the case for just this run or in general. For this these experiments were repeated 10,000 times. The results can be observed in table 4.1 and Figure 4.6 From the results in Table 4.1 one can conclude that on average the values of $\hat{b}_{intervals}$ are closer to $b = 1$ then the values of $\hat{b}$. If n increases the estimate gets better and the approximation of the MSE gets smaller. Also the behavior of $\hat{k}$ changes, it gets much smaller, which indicates that the error is noticed sooner. This can be explained by the fact that the values of $\hat{b}_k$ are more precise since they are the result of more data. Moreover, if there is more data the confidence intervals get smaller and there are more $\hat{b}_k$'s for small values of t (where the quadratic factor does not yet have a big impact).

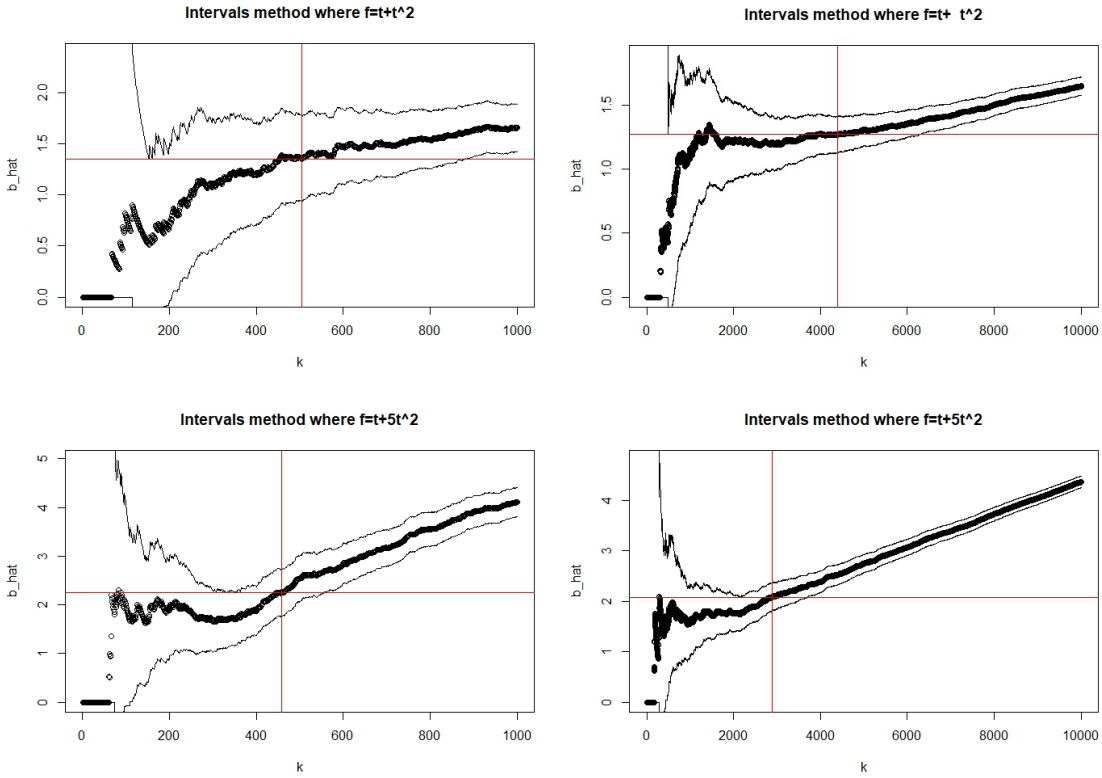
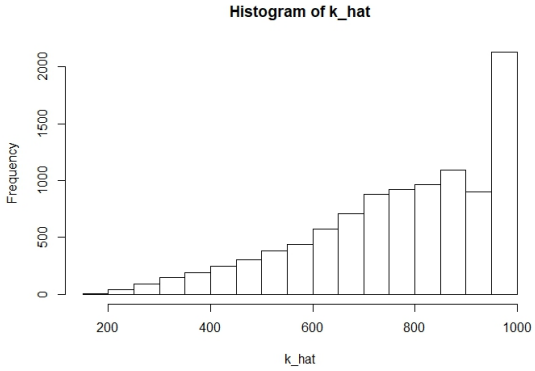
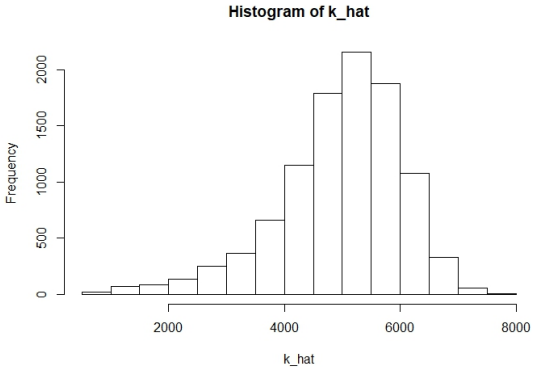


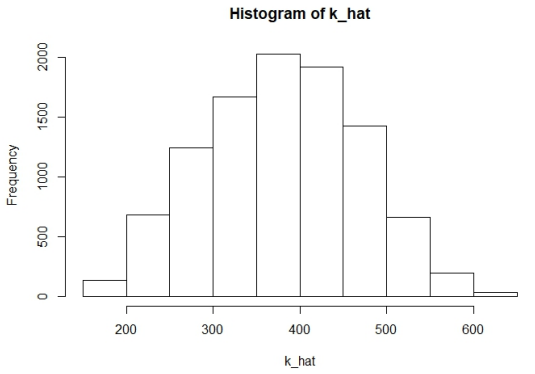
Figure 4.5: Behavior of method for quadratic f



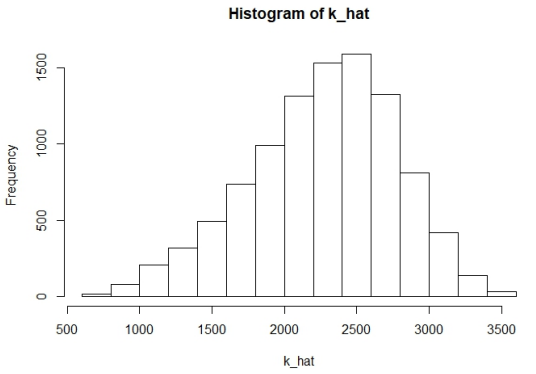
$f(t) = t + t^2$, $n = 1000$



$f(t) = t + t^2$, $n = 10000$

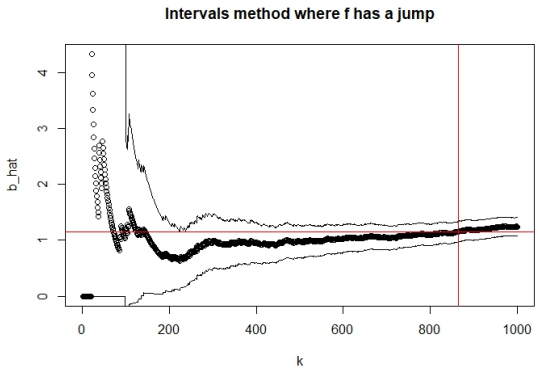


$f(t) = t + 5t^2$, $n = 1000$

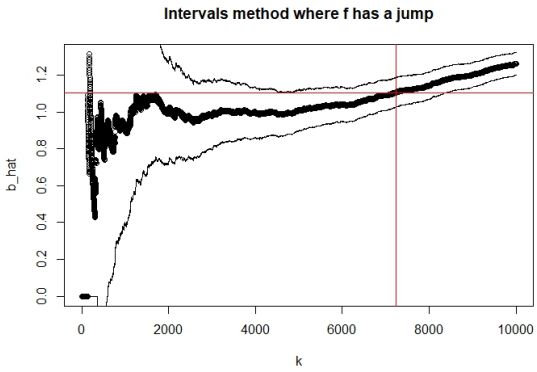


$f(t) = t + 5t^2$, $n = 10000$

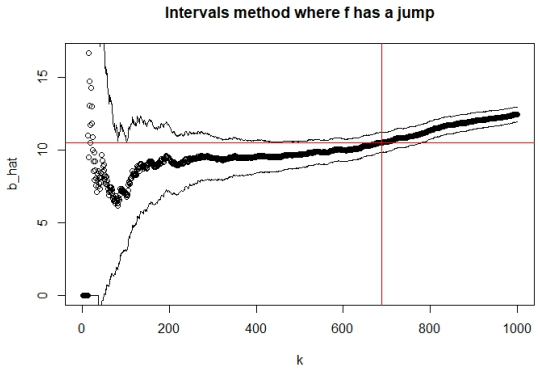
Figure 4.6: Histograms of $\hat{k}$



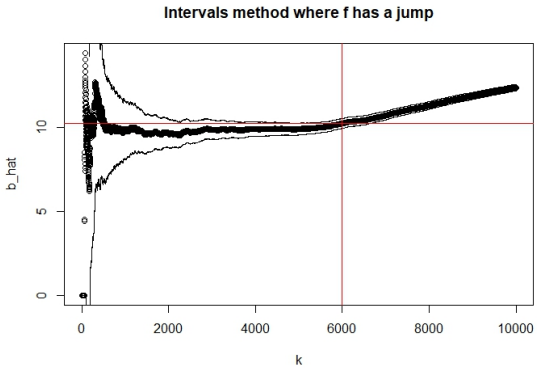
$b_0 = 1$, $b_1 = 2$, $t_0 = 0.5$ and $n = 1000$



$b_0 = 1$, $b_1 = 2$, $t_0 = 0.5$ and $n = 10000$



$b_0 = 10$, $b_1 = 20$, $t_0 = 0.5$ and $n = 1000$



$b_0 = 10$, $b_1 = 20$, $t_0 = 0.5$ and $n = 10000$

Figure 4.7: Behavior of method for f with a jump

If c increases and thus the quadratic impact increases $\hat{b}_{intervals}$ is still closer to 1 than $\hat{b}$. Moreover, $\hat{k}$ is smaller on average than when c is smaller. Finally, the approximated MSE increases if c increases since the deviation from the linear function is bigger. Thus, it can be concluded that the approximated MSE of the intervals method is lower than the approximated MSE of the $\hat{b}$ estimate in this case.

Case 3 f with a jump
This sort of function was already used to look at the quality of the bootstrap method in section 3.3.2. The function is defined as follows

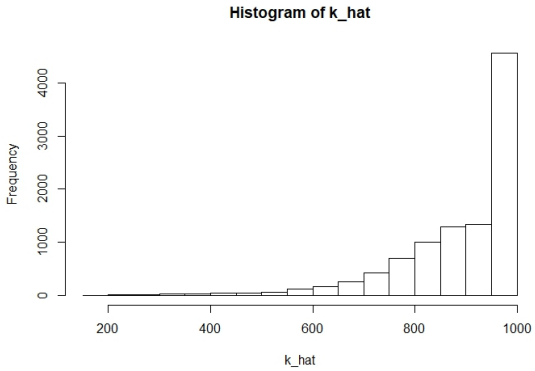
$$f(t) = \begin{cases} b_0 t & \text{if } t < t_0 \\ b_1(t - t_0) + b_0 t_0 & \text{if } t \geq t_0 \end{cases} .$$

In Figure 4.7 one can observe one run of the intervals method. Again the confidence intervals are constructed with $\alpha = \frac{0.05}{n}$ and the data comes from the Poisson distribution with parameter $f(t_i)$. It is clear that for these runs $\hat{k} < n$ and $\hat{b}_{intervals} < \hat{b}$.

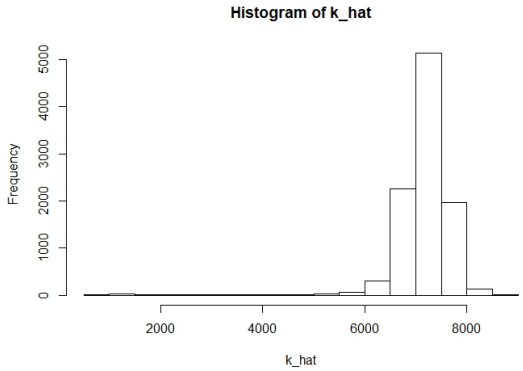
The question arises if this only happened by coincident or if this is always the case. Therefore, these experiments were repeated 10,000 times. The results can be observed in Table 4.2 and in Figure 4.8. Again it can be observed that the method works better if n is larger, the average estimate is better and the approximated MSE is smaller. In the case of $b_0 = 1, b_1 = 2$

b_0	b_1	n	mean $\hat{b}$	mean $\hat{b}_{intervals}$	mean $\hat{k} = n$	$\text{MSE}_{\hat{b}}$	$\text{MSE}_{\hat{b}_{intervals}}$
1	2	1000	1.2498	1.1837	0.3098	0.06491	0.04208
1	2	10000	1.2498	1.0890	0	0.06264	0.009199
10	20	1000	12.5036	10.6929	0	6.2923	0.6588
10	20	10000	12.5006	10.2967	0	6.2556	0.1124

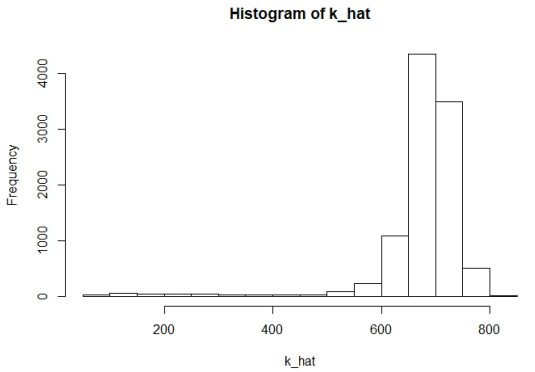
Table 4.2: Results from the intervals method for f with a jump



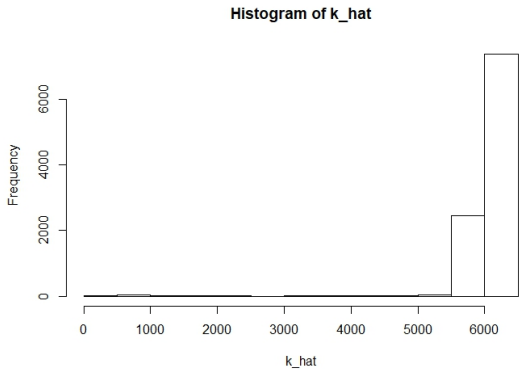
$b_0 = 1$, $b_1 = 2$, $t_0 = 0.5$ and $n = 1000$



$b_0 = 1$, $b_1 = 2$, $t_0 = 0.5$ and $n = 10000$

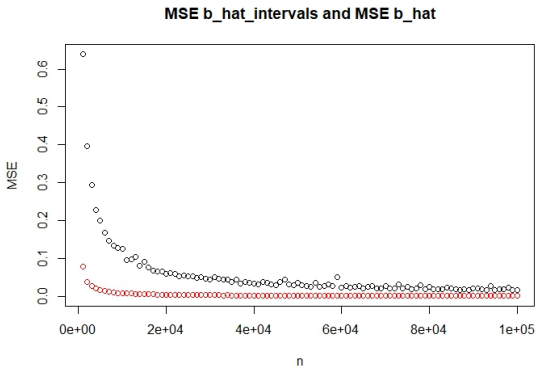


$b_0 = 10$, $b_1 = 20$, $t_0 = 0.5$ and $n = 1000$

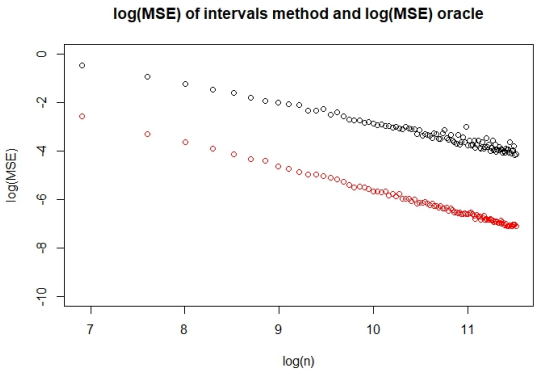


$b_0 = 10$, $b_1 = 20$, $t_0 = 0.5$ and $n = 10000$

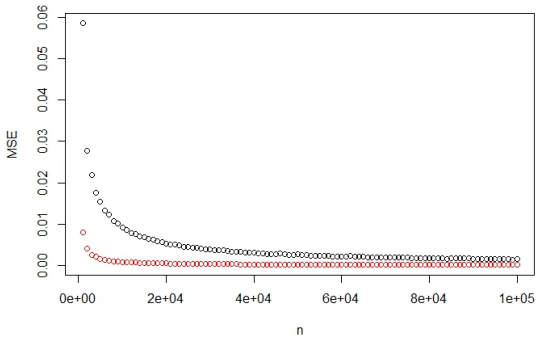
Figure 4.8: Histograms of $\hat{k}$ for f with a jump in the derivative



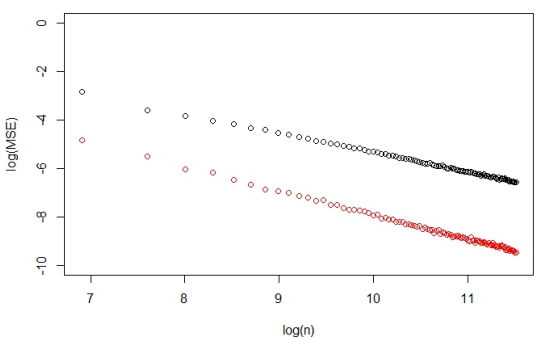
$b_0 = 10$, $b_1 = 20$, $t_0 = 0.5$



$b_0 = 10$, $b_1 = 20$, $t_0 = 0.5$



$b_0 = 1$, $b_1 = 2$, $t_0 = 0.5$



$b_0 = 1$, $b_1 = 2$, $t_0 = 0.5$

Figure 4.9: Comparison MSE intervals method and oracle

and $n = 1000$ in approximately 31% of the cases the method does not detect something strange (conflicting confidence intervals) and picks the very last $\hat{b}_k$. However, for a larger n this problem is solved.
In the histograms 4.8 it can be observed that for a larger value of n the value of $\hat{k}$ is much closer to $0.5 \cdot n$. This is of course desired since up and until this point $0.5 \cdot n$ the data comes from the linear function.

How well this intervals method works for this case can also be investigated by comparing it to $b_{0.5n}$. Since $b_{0.5n}$ is the estimator using all data where $f(t) = b_0 \cdot t$. This estimator will be called the oracle since this estimator knows until which point the function is linear (without a change in derivative). An approximation of the MSE is calculated for both estimators after 1000 repetitions for multiple values of n. The results are in Figure 4.9. Note that the red points are the mean squared errors of the oracle and the black ones of the intervals method estimator. It can be observed in Figure 4.9 that the mean squared errors decrease exponentially. Therefore, also the log-log scale of the mean squared errors and n are given. In that plot it can be observed that the points form a straight line with approximately the same derivative. This implies that the data fits a power law relation such as MSE$= a \cdot n^k$ and thus $\log(\text{MSE}) = k \cdot \log(n) + \log(a)$. Thus, the slope of the log-log plot gives the exponent k in this case. The value of a will be different which can be explained by the fact that the intervals method is not able to find the exact point where the derivative of f changes.

Overall, the intervals method shows potential. It is able to note if the function f is linear and

pick mainly the final $\hat{b}_k$. Furthermore, if f is not linear the method is able to notice this. However, this does not happen immediately. This implies that some data points are still taken into account while they might not be useful anymore. However, this might also be a good thing since otherwise the method would stop too fast in the case f is linear. The method could still be improved in some ways. This, however, falls beyond the scope of this bachelor final project. Some ideas are for example, looking at the confidence intervals (a tighter bound for the confidence intervals could make the method more precise) or making the method adaptive (choosing T_i around $\hat{k}$).

Chapter 5

Conclusions and discussion

This bachelor final project is motivated by a problem of testing quantum devices. Quantum devices contain elements which are unreliable. When doing an experiment one observes whether, after some time t, such an element is broken or not. The probability that it is broken is denoted by $f(t)$, which is a function that is assumed to be linear for small values of t. Two problems will be considered. Can we test if the function f is linear in the range of times we observed? And can we estimate the derivative of f at 0, with and without making further assumptions on this function?

In Chapter 2 an estimator of the derivative of f at 0 was derived under the assumption that f is linear. In Chapter 1 the quantity $a = f(0) = 0$ was set to 0. This was done, because this greatly simplified the estimator of b. The quantity a is approximately equal to 0, because the probability that an element breaks immediately after it has been installed is assumed to be very small. Therefore, in this report, a is assumed to be 0. Future research could investigate the case that a is actually a small positive number. Note that the estimator $\hat{b}$ will change and that moving away from this assumption will require some work.

In Chapter 3 the method that was used to test if the function f is linear was introduced. This consists of a bootstrap procedure. Testing if f is linear is basically testing if the data comes from a certain distribution (in this case Poisson) with the function $f(t)$ as parameter. This means that we were actually looking for a goodness-of-fit test. However, since the data is not identically distributed, the path of the bootstrap method was chosen. Note that the result of the bootstrap method highly depends on the number of data points (n), the estimation of the data ($\hat{b}$) and the number of bootstrap samples (K). Although the test seemed to be well calibrated, it was seen that empirically that it was not very powerful. Furthermore, and even more troublesome, it seemed sometimes observations from an alternative model lead to lower rejection rates than under the null model. We currently do not have a good explanation for this phenomenon, and it remains an interesting topic for future work.

Some experiments were performed with an estimator depending on the results of the bootstrap method. One could envision an estimator which makes use of the bootstrap testing procedure. One could conduct many different tests each using only observations until certain value of t. With the p-values one could then investigate until approximately which point $f(t)$ is linear. One could also use the p-values for a weighted estimate, where the values of (y_i, t_i) are weighted. Both experiments lead to very poor estimation procedures, which are possibly partly due to the poor performance of the bootstrap approach.

The intervals method as described in Chapter 4 gives an estimate of b, where no further assumptions on f are necessary. It is built upon the confidence intervals of b. These confidence intervals were constructed with a normal approximation and assuming $f(t)$ is linear. However, the central limit theorem states that the sum of n random variables converges in distribution to

a normal random variable for a large values of n. The problem that for small values of k the confidence intervals might not be very precise was solved by setting these confidence intervals automatically to very large intervals.

To assure that the confidence intervals are simultaneously valid, a trade-off was made. By Boole's law the $\mathbb{P}(\forall k \ : \ b \in [L_k, U_k]) \geq 1 - \alpha$ if $[L_k, U_k]$ is the α/n confidence intervals of b. This is a quite harsh lowerbound of this probability. It might lead to larger confidence intervals then necessary. Consequently, conflicting confidence intervals in case of a non-linear f might only be detected later. Another method to determine a larger lowerbound for $\mathbb{P}(\forall k \ : \ b \in [L_k, U_k])$ might be considered in future research. One could for example envision creating confidence intervals using Poisson tail bounds or other confidence intervals like described in, for example, [7].

The intervals method uses two parameters which need to be chosen before the method is used. These are the value of $\sum_{i=1}^{j} Y_i$ such that for all $i < j$ the confidence intervals are set to $[0, 1000]$ and the value of α. In this report the number of failures before the confidence intervals are taken into account is set to 5. In the experiments done this worked. The value of α was set to 0.05 as is customary within statistics. In future research the effects of these quantities should be investigated.

The intervals method described in this report was actually not the first idea. Another way to use these confidence intervals is to count for every $\hat{b}_k$ in how many confidence intervals it is contained. Then the estimate would simply be stated as the $\hat{b}_k$ which is in the most intervals. However, this method was replaced by the method described in Chapter 4. The reason for this was that the first method did not take into consideration that the confidence intervals in the beginning might be more reliable than the final ones, since the assumption is that f is approximately linear for small t. However, this implies that one could envision other methods to select an optimal $\hat{b}_k$. This could be explored in the future.

These estimators and the test explained in this report give some insight in how the behavior of the degrading elements can be investigated. This report can be viewed as a starting point of research into the testing for a certain degrading behavior of the qubits and gates in a quantum device. The explained methods should be explored further, to be applied to the true experiments. More knowledge about the failure probability will give more insight into the degrading behavior of these elements. This is necessary for the development of such an important future device that will address scientific challenges that current technology can not yet tackle.

Bibliography

[1] Quantum technology-popular science description. https://www.chalmers.se/en/news/Documents/quantum_technology_popdescr_171114_eng.pdf. Chalmers University of Technology. 5

[2] Felix Abramovich and Ya'acov Ritov. *Statistical theory : a concise introduction*. CRC Press,, 2013. ISBN: 9781439851845. 7, 8, 9, 10, 12, 14

[3] S. E. de Graaf, L. Faoro, L. B. Ioffe, S. Mahashabde, J. J. Burnett, T. Lindström, S. E. Kubatkin, A. V. Danilov, and A. Ya. Tzalenchuk. Two-level systems in superconducting quantum devices due to trapped quasiparticles. *Science Advances*, 6(51), 12 2020. DOI: 10.1126/sciadv.abc5055. 6

[4] John Loeffler. 5 Intractable Problems Quantum Computing Will Solve. https://interestingengineering.com/5-intractable-problems-quantum-computing-will-solve, 12 2018. 5

[5] National Academy of Engineering. *Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2018 Symposium*. National Academies Press, 1 2019. DOI: 10.17226/25333. 5

[6] Winfried Stute, Wenceslao Gonzáles Manteiga, and Manuel Presedo Quindimil. Bootstrap based goodness-of-fit-tests. *Metrika: International Journal for Theoretical and Applied Statistics*, 40(1), 1993. DOI: 10.1007/BF02613687. 13

[7] Manlika Tanusit. Two-Side Confidence Intervals for the Poisson Means . *International Journal of Modeling and Optimization*, 2(5), 10 2012. DOI: 10.7763/IJMO.2012.V2.189. 36

Appendix A

A.1 Extra figures

To compare the Bernoulli and Poisson model the $\mathbb{P}(Y_i \geq 1)$ in the case of the Poisson model and the $\mathbb{P}(Y_i = 1)$ in the case of the Bernoulli random variable are compared.

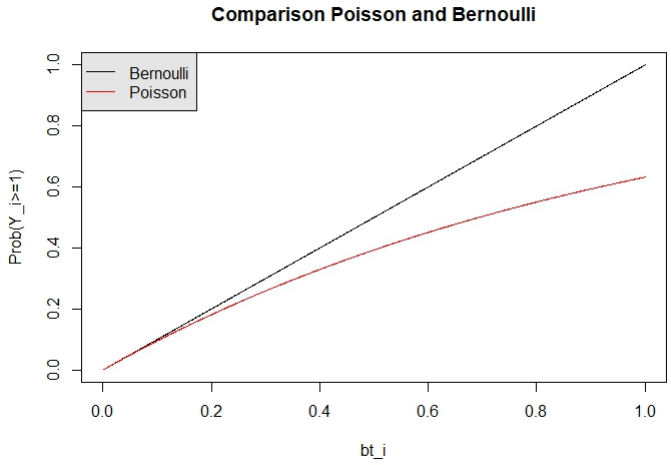


Figure A.1: Comparison probability $Y_i \geq 1$ Bernoulli and Poisson

In figure A.1 it can be observed that for small values of $b \cdot t_i$ the lines are very close together.

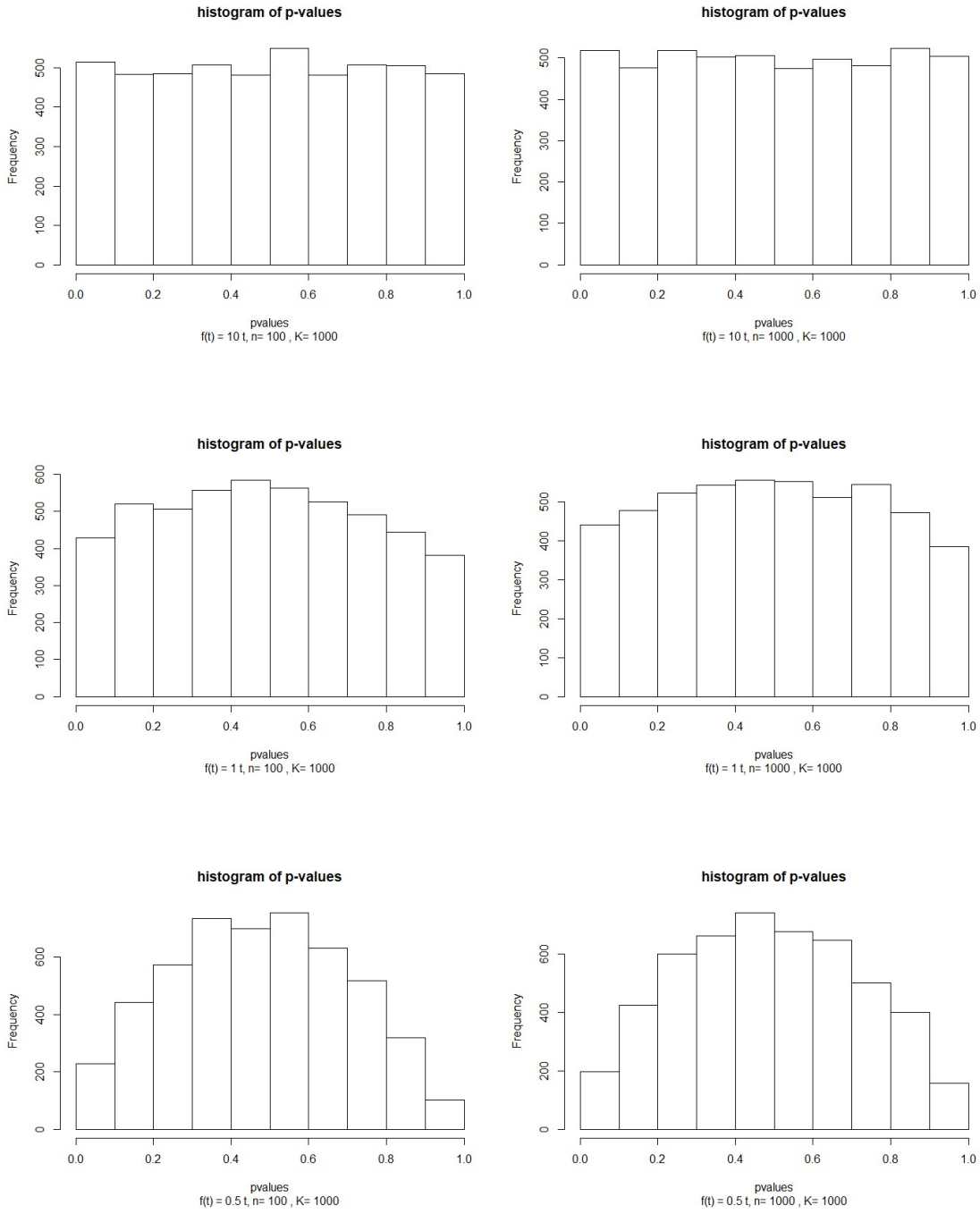


Figure A.2: Histograms for multiple values of b and n

A.2 R script Bootstrap method

```
b<-10
pvalues<-rep(0,5000)
n<-1000
```

```
#number of bootstrap interations
K<- 1000
for(l in 1:5000){
#Generate data
#original b
#b0 <- 1
# t0 <- 0.5
# b1 <- 2
# f <- function(t,b0,t0,b1){b0*t*(t<=t0)+(b1*(t-t0)+b0*t0)*(t>t0)}
T<- seq(1/n,1,by=1/n)
Y_original<- rpois(n,T+T^2)
#Y<-abs(rnorm(n,T,1))
#b_hat of data
#plot(T,Y, ylab='Y_i', xlab='t')
#lines(T,b_h*T, col=2)

b_h <- sum(Y_original)/sum(T)
#calculate test statistic
TS<- (sum(2*(b_h*T-Y_original)+2*Y_original*(log(Y_original+10^-323)-log(b_h*T
    +10^-323))))
# Bootstrap method
#create array to store test statistics
TSS<-rep(0,K)
for(k in 1: K){
  #create new data
  Y <- rpois(n,b_h*T)
  # Calculate new b_hat, on new data
  b_hat <- sum(Y)/sum(T)
  #calculate new test statistic
  TSS[k]<- (sum(2*(b_hat*T-Y)+2*Y*log((Y/(b_hat*T))+10^-120)))
}
#calculate p_value
pvalues[l]<- sum(TSS>TS)/K
}
hist(pvalues, main = 'histogram of p-values', sub=paste('f(t)=',b0,'t for t<',t0,'
    else f(t)=',b1,'(t-',t0,')+',b0*t0,", n=",n,', K=',K))
# proportion <0.05
mean(pvalues<0.05)
plot(sort(pvalues), main='sorted p-values', ylab='sorted p-values', sub=paste('f(t)
    =',b0,'t for t<',t0,'else f(t)=',b1,'(t-',t0,')+',b0*t0,", n=",n,', K=',K))
segments(x0=0,x1=5000,y0=0,y1=1, col=2)
prop[c]<-mean(pvalues<0.05)
prop
```

A.3 R Script Intervals method

```
f<-function(b,t){b*t}
g<-function(b,c,t){b*t+c*t^2}
h<-function(b0,b1,t0,t){b0*t*(t<=t0)+(b1*(t-t0)+b0*t0)*(t>t0)}

#intervals method
n<-1000
b<-1
c<-5
  T<- (1:n)/n
  alpha<-0.05/n
  z<-qnorm(1-(alpha/2))
  b_intervals<-rep(0,10000)
  b_hats<-rep(0,10000)
  k_list<-rep(0,10000)
for(i in 1:10000){
Y<-rpois(n,g(b,c,T))
b_h<-cumsum(Y)/cumsum(T)
lower<-b_h-z * sqrt(b_h/cumsum(T))
```

```
upper<-b_h+z * sqrt(b_h/cumsum(T))
lower<-lower*(cumsum(Y)>5)
upper<-upper*(cumsum(Y)>5)+1000*(cumsum(Y)<=5)
tmp<-(((cummin(upper)-b_h)>=0)*((b_h-cummax(lower))>=0))
b_intervals[i]<-b_h[n+1-which.max(tmp[n:1])]
k_list[i]<-n+1-which.max(tmp[n:1])
b_hats[i]<-b_h[n]
  }
mean(b_hats)
mean(b_intervals)
hist(k_list, main='Histogram of k_hat', xlab='k_hat')
mean(k_list==n)
mean((b_intervals-b)^2)
mean((b_hats-b)^2)
```