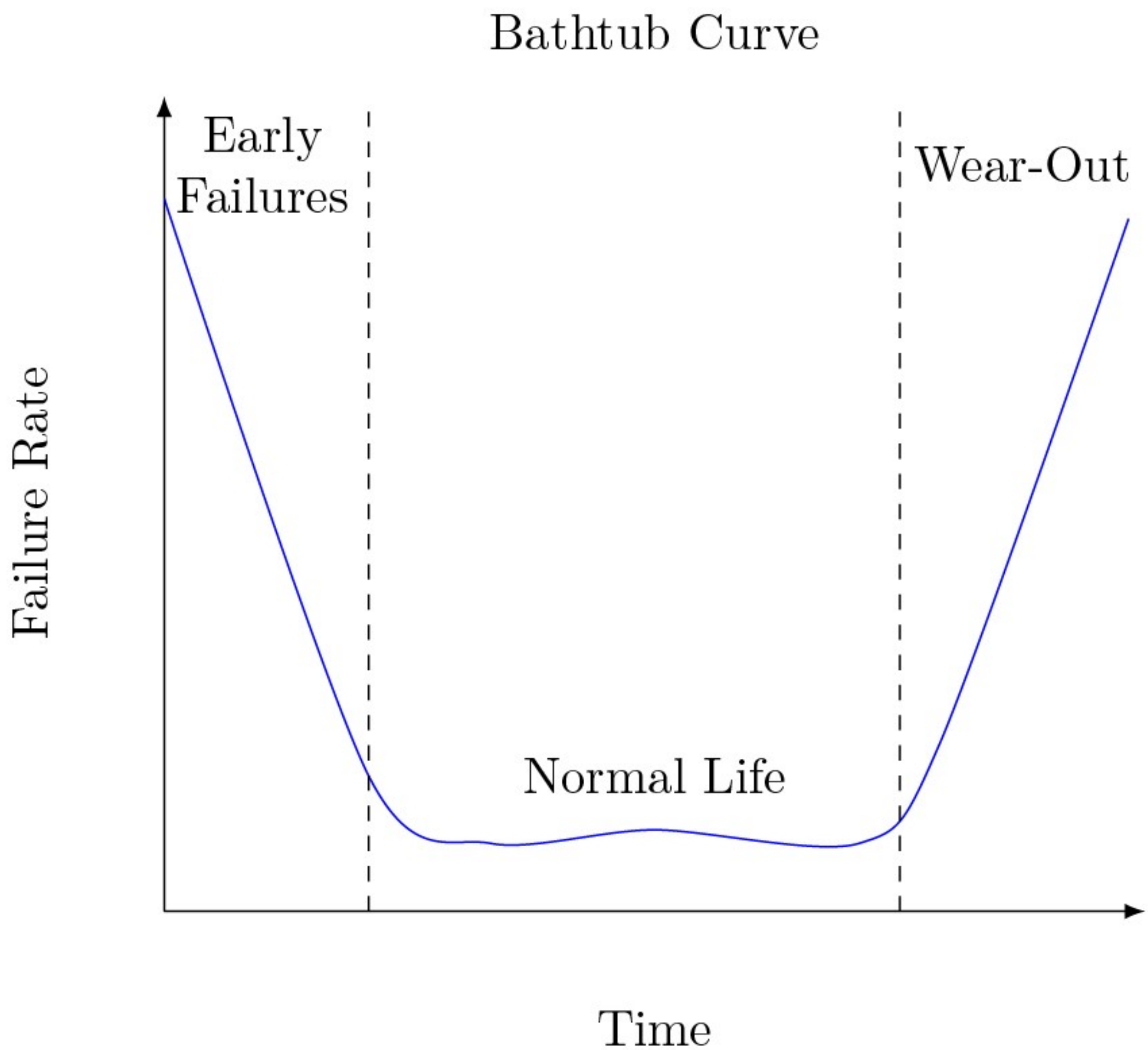
University of Stavanger

BENJAMIN HANSEN NORDHAUG
SUPERVISOR: JAN TERJE KVALØY

Test for trends in recurrent events data

Master thesis (2024)
Master of Science in Mathematics and Physics
Faculty of Science and Technology
Department of Mathematics and Physics



Test for trends in recurrent events data

Master thesis
University of Stavanger (UiS)

Benjamin Hansen Nordhaug

June 2024

Preface

I would like to thank my supervisor Jan Terje Kvaløy for his assistance and insights. His guidance, patience and feedback has helped me though this thesis.

Abstract

The aim of this thesis was to analyse how different trend tests perform under various scenarios, with a particular focus on simulated failure truncated data. The research conducted involved simulating failure truncated processes in R and analysing some time truncated real data. Key findings reveal that the choice of $\hat{\sigma}$ impacts the performance of the Lewis-Robinson and Anderson-Darling type tests, especially in scenarios with a low number of failures. The Laplace and Military Handbook tests were found unsuitable for Renewal Process (RP) and Trend-Renewal Process (TRP) situations, so they should be limited to scenarios where Poisson process is very highly likely. The Mann-Kendall test was observed to require a more pronounced trend, for increasing trends, when working with Weibull TRP situations. Furthermore, all the tests that were studied had challenges with high variance for less pronounced trends. This indicates a potential limitation of these tests in real-world scenarios where there are high variance.

Contents

List of Figures

Chapter 1

Introduction

Events that occur multiple times are called recurrent events, and the events do not need to be at regular intervals. Examples for recurrent events are disease outbreak, weather events, and traffic accidents.

A classic example of recurrent event data is repairable systems data. Repairable system usually fail repeatedly, and we can model the process. Here, we are interested in when the failures occur to see if there is a trend of when the failures occur. A repairable system like a car, can have several failures in the beginning. After the teething troubles are fixed, the problems should come less often. When time goes on, the system could become aging and have more failures. We call this trend a bathtub trend. Figure 1.1 shows how a bathtub model might look.

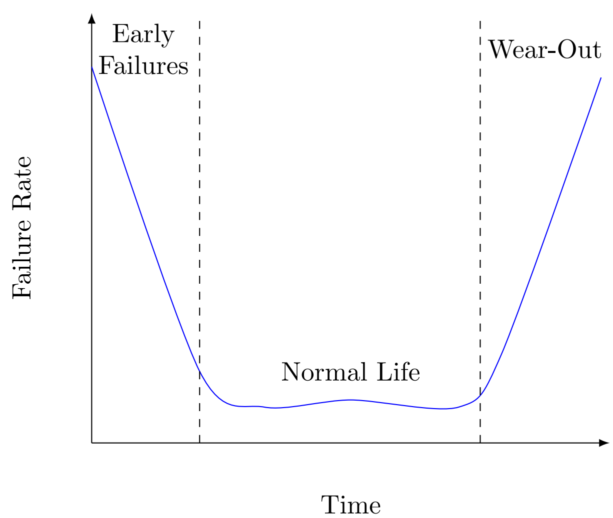


Figure 1.1: Bathtube Curve

Other types of trends is a decreasing trend, like the early failures part, of Figure 1.1. Then the opposite, is the increasing trend, which looks like the wear-out part of Figure 1.1.

Another cause for trends in data could be seasonal variation. Seasonal variation refer to changes in the frequency for the events for certain times of the year. Seasonal variation could for example be allergies, influenza, and extreme weather events.

This thesis will examine four different types of processes in Chapter 2, where especially the Trend-Renewal Process (TRP) will be relevant later in the thesis. Next, in Chapter 3, we will discuss various tests, including the Laplace, Mann-Kendall, and Anderson-Darling type tests. We want to use these tests to identify potential trends. In Chapter 4, we will explain how to simulate the different processes like the Trend-Renewal Process (TRP). Following that, in Chapter 5, we will look at the properties of the tests by simulating and analysing plots. Finally, in Chapter 6, we will apply some real data examples to see how we can use the tests to analyse data.

Chapter 2

Introduction to type of processes

We start by looking at notation and truncation. Figure 2.1 shows events that occur at event time points $\{T_1, T_2, T_3, ...\}$. These events could, e.g. be failures in a system. Further, $\{X_1, X_2, X_3, ...\}$ are the times between these events, we often call these interarrival times. The number of failures in an interval $[0, t]$, is written as $N(t)$. Figure 2.1 shows that we stop looking at failures after a

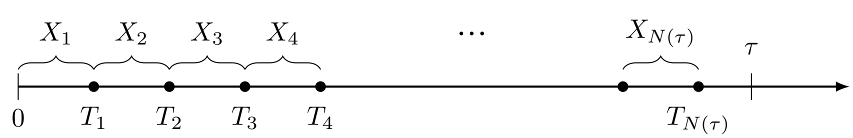


Figure 2.1: Time truncated process

predefined time τ. That is called a time truncated process. That means that the number of failures is a random variable, $N(\tau)$.

Another censoring scheme for recurrent events is a failure truncated process. With a failure truncated process, we stop observing after a predefined number of failures, this means that the process ends up at a failure. The time the observation stops is now a random variable T_n. Figure 2.2 shows what a failure truncated process may look like.

Figure 2.2: Failure truncated process

The rate of occurrence of failures (ROCOF) is a measure of the rate of the failure times, and can be written as $w(t) = \frac{dE(N(\tau)}{dt}$[1, p.150]. On the other hand, the intensity is the expected rate which the events occur, given the history of the process. The intensity function is written as Equation 2.1.

$$\phi(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\left(N\left(t + \Delta t\right) - N(t) = 1 | \mathcal{F}_{t-}\right) \tag{2.1}$$

In Equation 2.1, the term $\mathcal{F}_{t-}$ is the history of the process up to right before the time t. Firstly, we start by introducing four basic stochastic point processes which are useful for modeling repairable systems [2, p.29].

2.1 Homogeneous Poisson Process (HPP)

We begin with a type of point process known as the homogeneous Poisson process. A simple homogeneous Poisson process may look like Figure 2.1. The conditions for HPP are as follows.

1. $N(0) = 0$
2. $N(a) - N(b)$ is independent of $N(c) - N(d)$ when $[b, a]$ and $[d, c]$ are non-overlapping intervals.
3. $P(N(t + \Delta t) - N(t) = 1) \approx \lambda \cdot \Delta t$
4. $P(N(t + \Delta t) - N(t) > 1) \approx 0$

It can be proven from these conditions, that for the process to be HPP, $\{X_1, X_2, X_3, ...\}$ must be identically exponentially distributed with parameter λ and independent. In the list, the first condition for HPP is $N(0) = 0$. This means that counting of events start at $t = 0$.

From the third condition in the list, we can see that the process is using the parameter λ. The ROCOF and the intensity in HPP is λ, so it is constant for HPP. The rate in HPP does not change with time, which is the homogeneous part of the HPP. Another thing is that $\Lambda(t) = \lambda t$ is the expected number of events by time t.

Finding the probability of observing a certain number of events in a given time interval $(t, t + s]$ can be calculated with Equation 2.2. This is a Poisson distribution, and the probability is the same for all t.

$$P\left((N(t + s) - N(t)) = j\right) = \frac{e^{-\lambda s}(\lambda s)^j}{j!} \tag{2.2}$$

A process with HPP cannot have a trend, as the intensity of failures does not change over time.

2.2 Non-Homogeneous Poisson Process (NHPP)

Unlike for a homogeneous Poisson process, a non-homogeneous Poisson process does not have a constant intensity λ. For an NHPP, the λ changes with time, so we use $\lambda(t)$. Based on the same conditions as for HPP, we then get these conditions for an NHPP.

1. $N(0) = 0$
2. $N(a) - N(b)$ is independent of $N(c) - N(d)$ when $[b, a]$ and $[d, c]$ are non-overlapping intervals.
3. $P(N(t + \Delta t) - N(t) = 1) \approx \lambda(t) \cdot \Delta t$
4. $P(N(t + \Delta t) - N(t) > 1) \approx 0$

When we compare these four conditions with the HPP, the only thing that has changed is the third point. While with an HPP the probability of observing a certain number of events in an interval $(t, t + s]$ does not change with time, with an NHPP it does. Using these conditions, the probability of observing a certain number of events in a given time interval for an NHPP model can be calculated with Equation 2.3[1, p.449][3, p.102]. Like HPP, this is still a Poisson distribution, but now depends on t.

$$P\left((N(t+s) - N(t) = j\right) = \frac{e^{-\int_t^{t+s} \lambda(v)dv} \left\{\int_t^{t+s} \lambda(v)dv\right\}^j}{j!} \tag{2.3}$$

We can see from Equation 2.3, that if we want to calculate the expected number of events in the interval $(t, t + s]$, we can use Equation 2.4.

$$E\left[(N(t+s) - N(t)\right] = \int_t^{t+s} \lambda(v)dv \tag{2.4}$$

A special case of Equation 2.4 to look at, is when we get the mean value function [3, p.102] for an NHPP, given as Equation 2.5.

$$E(N(t)) = \Lambda(t) = \int_0^t \lambda(u)du \tag{2.5}$$

Equation 2.5, which shows the expected number of failures in the interval $(0, t]$, is a special case of Equation 2.4. To illustrate this we can look again at the bathtube curve as in Figure 2.3.

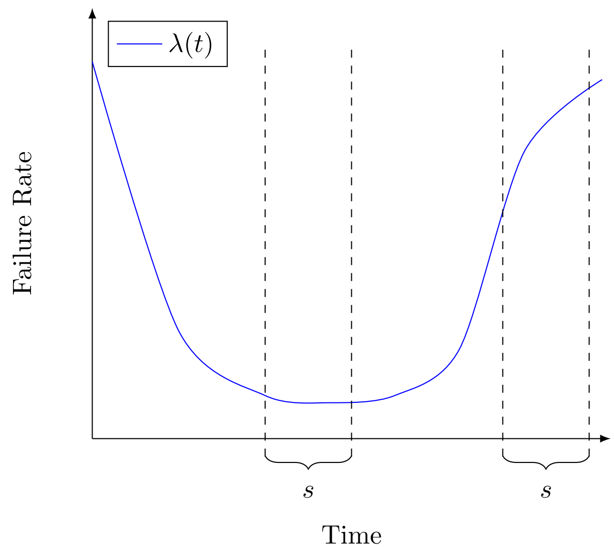


Figure 2.3: Bathtube Curve intensity

In Figure 2.3, we see that the intensity changes with time. For the first interval s, the failure rate is low, but for the second interval s, the failure rate is higher.

Unlike a process with an HPP, for an NHPP the intensity of failures can change with time. We saw that with $\lambda(t)$. This means that a process with an NHPP can have a trend.

2.3 Renewal Process (RP)

While the interarrival times $\{X_1, X_2, X_3, ...\}$ for an HPP are exponentially distributed, for a renewal process, the interarrival times can be independent and identically distributed with any continuous distribution F with support on $\mathbb{R}^+$. This makes the renewal process a generalisation of the HPP.

Since a RP is just an HPP with more choices for the distribution, the failure intensity does not change systematically with time. This means that a process with a RP cannot have a trend.

The intensity of a RP is given by Equation 2.6, where the interarrival times $\{X_1, X_2, X_3, ...\}$ are i.i.d. with respect to the distribution F [4].

$$\phi(t) = z\left(t - T_{N(t-)}\right) \tag{2.6}$$

The component $z()$ represents the hazard function corresponding to the distribution F. From this, we can observe that the intensities restart from z(0) at each event time, thereby indicating that there is no trend.

2.4 Trend-Renewal Process (TRP)

With NHPP being a generalisation of the HPP, and in a similar manner, TRP being a generalization of the RP. TRP is also a generalisation of NHPP. This relationship can be viewed as Figure 2.4 shows [4, p.35]. The intensity for a TRP is given by Equation 2.7 [4, p.33].

$$\phi(t) = z\left(\Lambda(t) - \Lambda\left(T_{N(t-)}\right)\right) \cdot \lambda\left(t\right) \tag{2.7}$$

In Equation 2.7, $\lambda(t)$ is the trend function and $\Lambda(t) = \int_0^t \lambda(v)dv$. An important fact is when $\lambda(t)$ is constant, then TRP simplifies to a RP.

The other component of Equation 2.7 is $z(t)$. An important fact is that when $z(t)$ is constant, TRP simplifies to an NHPP. This simplification, along with when $\lambda(t)$ is constant, is illustrated in Figure 2.4. It is important to note that TRP can have a trend, as long as $\lambda(t)$ is not constant. For more in-depth explanation of TRP, I refer to the work by Lindqvist et al. [4].

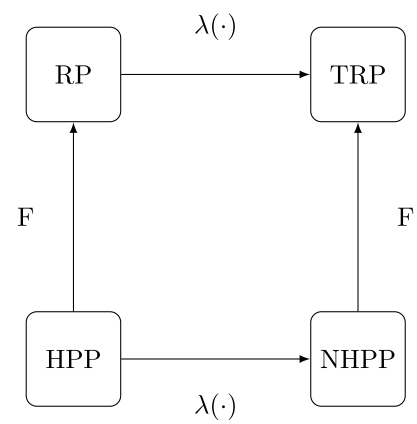


Figure 2.4: Illustration of the connections between the four processes

Let us show an example of TRP. Here we have $F \sim$ Weibull. The Weibull distribution is given by the density function in Equation 2.8.

$$f(t) = \alpha\beta t^{\beta-1}e^{-\alpha t^{\beta}} \tag{2.8}$$

The hazard function is given by Equation 2.9.

$$z(t) = \frac{f(t)}{1 - F(t)} \tag{2.9}$$

Using Equation 2.9, we get the hazard function for a Weibull distribution to be like Equation 2.10.

$$z(t) = \alpha\beta t^{\beta-1} \tag{2.10}$$

The trend function can be set as $\lambda(t) = abt^{b-1}$, which means we get $\Lambda(t) = \int_0^t \lambda(v)dv = at^b$. Using the trend function and Equation 2.10, to plug into Equation 2.7, we obtain the TRP intensity for a Weibull distribution, which is then given by Equation 2.11.

$$\phi(t) = \alpha\beta \left(at^b - aT^b_{N(t-)}\right)^{\beta-1} abt^{b-1} \tag{2.11}$$

It is now interesting to look back at Figure 2.4 to look at some special cases to examine for Equation 2.11. The first case is when $b = 1$, then the intensity is given by Equation 2.12.

$$\phi(t) = \alpha\beta \left(at - aT_{N(t-)}\right)^{\beta-1} a = \alpha^*\beta \left(t - T_{N(t-)}\right)^{\beta-1} \tag{2.12}$$

where $\alpha^* = a \cdot \alpha$. Looking at Equation 2.12, it now describes the RP intensity for a Weibull distribution. Another case to look at for Equation 2.11, is when $\beta = 1$.

$$\phi(t) = \alpha \cdot 1 \cdot abt^{b-1} = \alpha^* bt^{b-1} = \lambda(t) \tag{2.13}$$

Equation 2.13 now describes the NHPP intensity for a Weibull distribution. The last special case is then, when both $b = 1$ and $\beta = 1$, which is calculated in Equation 2.14.

$$\phi(t) = \alpha^* \cdot 1 \cdot t^{1-1} = \alpha^* \tag{2.14}$$

As seen in Equation 2.14, when both $b = 1$ and $\beta = 1$, it is describing an HPP intensity.

Chapter 3

Tests for trends

This chapter will explain five trend testing methods. The first two tests, the Laplace and Military Handbook, utilise the null hypothesis of HPP.

3.1 Laplace test

The Laplace test is a test which checks if a single system has a trend or not. The null hypothesis H_0, is that the process is HPP. The alternative hypothesis H_1 is that the process is not HPP. Based on how the test works, we expect a test with power against monotonic trends. The formula for a Laplace test with time truncated data is given by Equation 3.1.

$$L_\tau = \frac{\sqrt{12}}{\tau \cdot \sqrt{N(\tau)}} \cdot \left[\sum_{i=1}^{N(\tau)} T_i - \frac{N(\tau)}{2} \tau \right] \tag{3.1}$$

Equation 3.1 shows the formula when we have a time truncated process, while Equation 3.2 shows the formula for a failure truncated process.

$$L_n = \frac{\sqrt{12}}{T_n \sqrt{n-1}} \cdot \left[\sum_{i=1}^{n-1} T_i - \frac{n-1}{2} T_n \right] \tag{3.2}$$

We let L be either L_τ or L_n, depending on whether the process is a time or failure truncated process. We find that L is approximately normally distributed [5, p.16]. Before concluding based on a calculated value, L from Equation 3.1 or Equation 3.2, we have to choose the level of test to use. If we use a 5% level on the test, and use a two-sided test, then the rejection limit is ± 1.96. After we have calculated L, we can check it against this limit or calculate a p-value. If the value is greater than 1.96 or less than -1.96, we can discard the null hypothesis. Another thing about the Laplace test is that the sign of L indicates the direction of the trend. When $L > 0$, the trend increases, which means that it is a degradation trend. For the opposite case, where $L < 0$, the trend decreases, which means that it is an improvement trend. We can see this intuitively from the formula for the Laplace test. If we look at the time truncated case, the term $\sum_{i=1}^{n-1} T_i - \frac{n-1}{2} T_n$, shows us that many large T_i and few small T_i will output a positive number. Remember that T_i is the time when the event occurs, counting from when the process begins. Thus, a positive number for the term will mean that $L > 0$. Conversely, when the process has many small T_i and few large T_i, we then get $L < 0$.

3.2 Military Handbook test

The Military Handbook test is similar to the Laplace test, but the distribution of the test statistic is a chi-square distribution. The null hypothesis H_0, is where the process is an HPP. The alternative hypothesis H_1 is that the process is not an HPP. It can be shown that the following test statistic will have a chi-square distribution with $2n$ degrees of freedom [5, p.17].

$$MH_\tau = 2 \sum_{i=1}^{n} \ln \frac{\tau}{T_i} \tag{3.3}$$

As in the Laplace test, Equation 3.3 shows the formula when we have a time truncated process, while Equation 3.4 shows the formula for a failure truncated process. So, it can be shown that Equation 3.4 will have a chi-square distribution with $2(n-1)$ degrees of freedom [5, p.17].

$$MH_n = 2 \sum_{i=1}^{n-1} \ln \frac{T_n}{T_i} \tag{3.4}$$

Depending on if we use Equation 3.3 or Equation 3.4, let MH be either MH_τ or MH_n. When calculating MH (either Equation 3.3 or Equation 3.4), the next step is to compare it to the percentiles of the chi-square distribution with the appropriate degrees of freedom. If we want to check if there is a trend, we can use a two-sided test. If we use a 5% level on the test, then we use 2.5% on each side to compare with.

It is also possible to do a one sided test, to only check for an increasing trend, or for a decreasing trend. When checking if there is an increasing trend, we only check against the left side of the chi-square distribution. Accordingly, to check for a decreasing trend, we can check against the right side of the chi-square distribution.

3.3 Lewis-Robinson test

The Lewis-Robinson test is very similar to the Laplace test, but instead of using HPP as the null hypothesis H_0, the Lewis-Robinson test is constructed to have RP as H_0. For the alternative hypothesis H_1, the process is not RP. As with the Laplace test and the Military Handbook, we have both a formula for time truncated process and one formula for failure truncated process. The formula for time truncated process is given by Equation 3.5, as detailed in the work by Kvaløy & Lindqvist [6].

$$LR_\tau = \frac{\bar{X}}{\hat{\sigma}} \cdot \frac{\sqrt{12}}{\tau \cdot \sqrt{N(\tau)}} \left[\sum_{i=1}^{N(\tau)} T_i - \frac{N(\tau)}{2}\tau \right] \tag{3.5}$$

Equation 3.5 is for time truncated process, while Equation 3.6 is for a failure truncated process.

$$LR_n = \frac{\bar{X}}{\hat{\sigma}} \cdot \frac{\sqrt{12}}{T_n\sqrt{n-1}} \cdot \left[\sum_{i=1}^{n-1} T_i - \frac{n-1}{2}T_n \right] \tag{3.6}$$

Equation 3.6 is a variant of the original test from Lewis & Robinson [7]. An alternative derivation of the Lewis-Robinson test from Kvaløy & Lindqvist [8] is given by Equation 3.7.

$$LR_n^* = \frac{\bar{X}}{\hat{\sigma}} \cdot \frac{\sqrt{12}}{T_n\sqrt{n}} \cdot \left[\sum_{i=1}^{n-1} T_i - \frac{n-1}{2}T_n \right] \tag{3.7}$$

The test is used in the same manner as the Laplace test. The Lewis-Robinson test is approximately normally distributed [6, p.103]. As with the Laplace test, the rejection limit is ± 1.96, for a 5% level on the test.

We can see that the Lewis-Robinson Equations 3.5 and 3.6, are very similar to the Laplace Equations 3.1 and 3.2. However, for Lewis-Robinson the term $\frac{\bar{X}}{\hat{\sigma}}$ is added. This term $\frac{\bar{X}}{\hat{\sigma}}$ is called coefficient of variation and for HPP the term is approximately equal to 1. For RP the term differs from 1.

It is possible to use different standard deviation estimates, denoted as $\hat{\sigma}$, when calculating the test. In this context, we will denote the standard deviation using the typical method as $\hat{\sigma}_1$, which is calculated as the square root of the sample variance. $\hat{\sigma}_1$ is given by Equation 3.8.

$$\hat{\sigma}_1 = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2} \tag{3.8}$$

Equation 3.8 is for a time truncated process, while Equation 3.9 is for failure truncated process.

$$\hat{\sigma}_1 = \sqrt{\frac{1}{N(\tau)-1} \sum_{i=1}^{N(\tau)} (X_i - \bar{X})^2} \tag{3.9}$$

The other method proposed by [9, p.155] is denoted as $\hat{\sigma}_2$, which is given by Equation 3.10 or Equation 3.11.

$$\hat{\sigma}_2 = \sqrt{\frac{\sum_{i=1}^{n-1} (X_{i+1} - X_i)^2}{2(n-1)}} \tag{3.10}$$

Equation 3.10 is for a time truncated process, while Equation 3.11 is for failure truncated process.

$$\hat{\sigma}_2 = \sqrt{\frac{\sum_{i=1}^{N(\tau)-1} \left(X_{i+1} - X_i\right)^2}{2\left(N\left(\tau\right) - 1\right)}} \tag{3.11}$$

Equations 3.10 and 3.11 are variants that can be found explained in more detail in [6, p.104].

3.4 Mann-Kendall test

The Mann-Kendall test is another test, but this particular test is for failure truncated process only [10][11, p.165][12, p.208]. The Mann-Kendall test is based on finding the number of times that the interarrival time $\{X_1, X_2, X_3, ...\}$ is shorter than the previous interarrival time. This is also called counting the number of reverse arrangements. The Mann-Kendall test is given by Equation 3.12.

$$M = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I\left(X_i < X_j\right) \tag{3.12}$$

In Equation 3.12, we have that $I(A) = 0$ or 1 depending on whether the event A does not occur or does occur, respectively. If there are many reverse arrangements, then there is a sign of decreasing trend. On the other hand, if there are few reverse arrangements, then there is a sign of increasing trend.

When $n < 10$, you can find tables for the distribution for M. However, for $n \geq 10$, we have approximately normally distributed M for the null hypothesis. The expectation is then $n(n-1)/4$ and the variance is given by $(2n^3 + 3n^2 - 5n)/72$ [10, p.248-249].

Because we are counting the number of reverse arrangements, there is no specific distribution assumption for the process, so under H_0 the process can be any RP.

3.5 Anderson-Darling type test

The last test we are looking at in this chapter is the Anderson-Darling type test, which has a longer formula than the other tests we have previously seen. The test uses RP as the null hypothesis H_0. The alternative hypothesis H_1, is not RP. For a time truncated process, the Anderson-Darling test statistic is given by Equation 3.13[6, p.103-104].

$$AD_\tau = \left(\frac{\bar{X}}{\hat{\sigma}}\right)^2 \cdot \frac{1}{N(\tau)} \left\{ \sum_{i=1}^{N(\tau)-1} \left[(N(\tau)-i)^2 \ln\left(\frac{\tau - T_i}{\tau - T_{i+1}}\right) + i^2 \ln\left(\frac{T_{i+1}}{T_i}\right) \right] \right. \\ \left. + N(\tau)^2 \left[\ln\left(\frac{\tau}{\tau - T_1}\right) + \ln\left(\frac{\tau}{T_{N(\tau)}}\right) - 1 \right] \right\} \quad (3.13)$$

Subsequently, the formula for a failure truncated process is given by Equation 3.14 [4, p.4-5].

$$AD_n = \frac{n\bar{X}^2}{\hat{\sigma}^2} \sum_{i=1}^{n} \left[q_i^2 \ln\left(\frac{i}{i-1}\right) + (q_i + r_i)^2 \ln\left(\frac{n-i+1}{n-1}\right) - \frac{r_i^2}{n} \right] \quad (3.14)$$

where $q_i = (T_i - iX_i)/T_n$ and $r_i = nX_i/T_n - 1$.

We let AD be either AD_τ or AD_n, depending on whether the process is a time or failure truncated process. The Anderson-Darling test only discards the null hypothesis with high values of AD. To find the value to reject, we must use the Anderson-Darling distribution. Therefore, when using a 5% level on the test, the rejection value is 2.492[6, p.104]. An advantage with the Anderson-Darling test is that it has strength against both monotonic and non-monotonic trends[6, p.113].

To calculate the standard deviation denoted as $\hat{\sigma}$, we follow the same procedure as in the Lewis-Robinson test and use Equation 3.10 or 3.11.

Chapter 4

Simulation methods

In this chapter, we are going to look at the ways in which we can simulate the processes. The explanations are for failure truncated processes, but for time truncated processes we should simulate more interarrival times than needed, and then there should be a function to stop the process at the predefined time τ. If not, we might end up with a process that ends before the predefined time τ.

4.1 HPP

HPP can be simulated by generating the interarrival times $\{X_1, X_2, X_3, ...\}$ from the exponential distribution with a rate equal to a constant λ. Then we can change these interarrival times to event time points $\{T_1, T_2, T_3, ...\}$. This is done by simply calculating $T_j = \sum_{i=1}^{j} X_i$.

4.2 NHPP

For an NHPP we will consider two methods for simulating, which we will look at. These are inversion and thinning.

4.2.1 NHPP Inversion

When simulating an NHPP, we can start by using the same method as for HPP. Then we must transform the times using an inverse function, as Figure 4.1 shows. The inverse function is the inverse of the integrated intensity function (Λ), which is given by Equation 2.5.

It can be shown that by simulating an HPP with a rate equal to 1, and then transforming the simulated event times by the inverse Λ-function, we get an NHPP with intensity function λ [1, p.451-452].

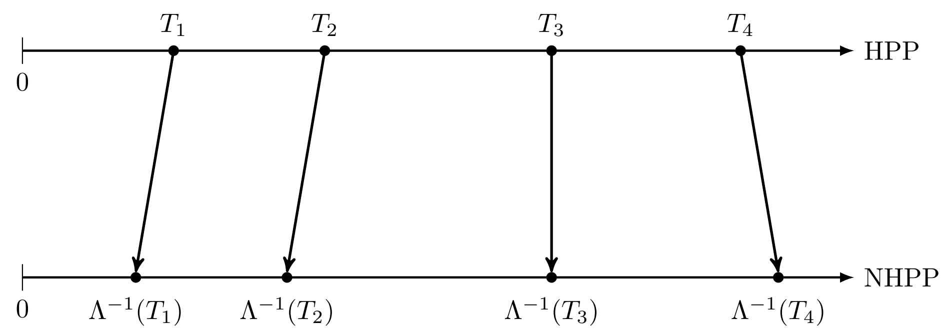


Figure 4.1: Transformation from HPP to NHPP.

4.2.2 NHPP Thinning

Another way to simulate NHPP is by using thinning. Thinning starts by simulating HPP and selectively discarding some events based on the desired rate function. Read more about thinning in Rizzo[3, p.102-104].

An advantage of the thinning method is that we do not need to invert the $\Lambda(t)$-function. This means that it is easier to use thinning for a lot of $\lambda(t)$-functions, than using inversion. Another advantage for thinning is that we do not need to a have a check for not running out of interarrival times when dealing with a time truncated process. However, with thinning, we need this check for failure truncated processes.

4.3 RP

RP can be simulated by generating the interarrival times $\{X_1, X_2, X_3, ...\}$ from any distribution with support on $\mathbb{R}^+$. Then we can change these interarrival times to time points $\{T_1, T_2, T_3, ...\}$. This is the same method as for HPP, but here we can use any distribution.

4.4 TRP

To simulate a TRP, we start by simulating the time points $\{T_1, T_2, T_3, ...\}$ from a RP. Using the inversion method, we can apply the inverse Λ-function of the trend function $\lambda(t)$. A visual representation of this method is shown in Figure 4.2. This is using the trick as simulating from HPP and transforming them to NHPP. This can be visually shown in Figure 2.4.

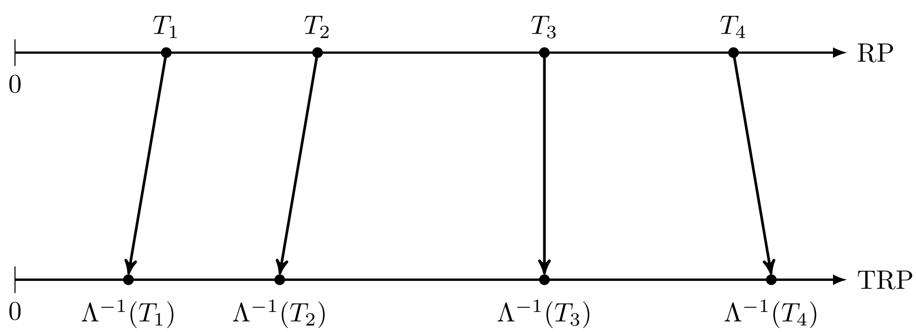


Figure 4.2: Transformation from RP to TRP.

Chapter 5

Simulation Study

In this chapter we are going to look at properties of the tests in various situations. For each choice of model and parameter values, we have used 20 000 simulations. A 5% level on the tests in all the plots in this chapter. The plots that display the number of failures on the x-axis use a step size of 1, simulating from 5 to 60 failures. The step size for the other plots is 0.01, with the value of the parameter changed in the simulation ranging from 0.01 to either 3.0 or 3.5. The Lewis-Robinson test used in these simulations is a variant of the original test, the one from Lewis & Robinson (3.6). For most of the plots that are not using an HPP, β will be discussed. In these plots, β is the shape parameter using a Weibull distribution. The plots in this chapter are simulated as failure truncated process, which simplifies the coding process and allows us to apply the Mann-Kendall test.

5.1 Level Properties

First, we start by looking at the tests level properties, in different situations without a trend. With no trend, ideally the tests should have a rejection probability at 5%, given that we used a 5% level on the tests. That is why there is a dotted line at 0.05 in all the plots that is looking at the level properties.

To begin with, we start with simulating HPP. Figure 5.1 shows how the different tests perform with different numbers of failures for an HPP.

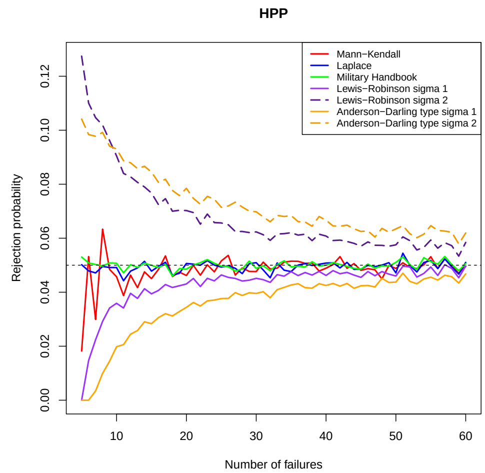


Figure 5.1: Level of the tests simulated as a function of the number of events. Data simulated from HPP with rate $\lambda = 0.01$. A nominal level of 0.05 was used.

As shown in Figure 5.1, the Mann-Kendall, Laplace, and Military Handbook tests mostly maintain a value close to 0.05 across all observed failure counts. However, the Mann-Kendall test struggles with very low number of failures. The Lewis-Robinson and Anderson-Darling tests behave quite similarly. For $\hat{\sigma}_1$ (3.9), their values are a bit low for the lower number of failures. They do get quite close for the higher number of failures. When we use $\hat{\sigma}_2$ (3.11) in the Lewis-Robinson and Anderson-Darling tests, both have a too high rejection probability, especially with a low number of failures. This means that the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_2$ are rejecting H_0 too often, which means they detect trends more often than they should. This means that it is more likely for the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_2$ to make a type I error for an HPP process [13, p.342]. Likewise, for the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_1$, a low number of failures makes them more likely than the other tests to make a type II error.

In context of the Weibull RP, the parameter β is important to understand the variance in the process. Overdispersion and underdispersion are two concepts that are used in this analysis. Overdispersion refers to the fact that β is lower than 1, which means that there is more variance in the process than expected under a Poisson process. In contrast, underdispersion refers to when β is higher than 1, which means that there is less variance in the process than expected in a Poisson process.

The next plots, Figure 5.2, show how the different tests perform with different numbers of failures for a RP. In Figure 5.2, all tests, except for the Military Handbook and Laplace tests, behave mostly the same as in Figure 5.1. In the left plot, the Military Handbook and Laplace tests, have a way too high rejection

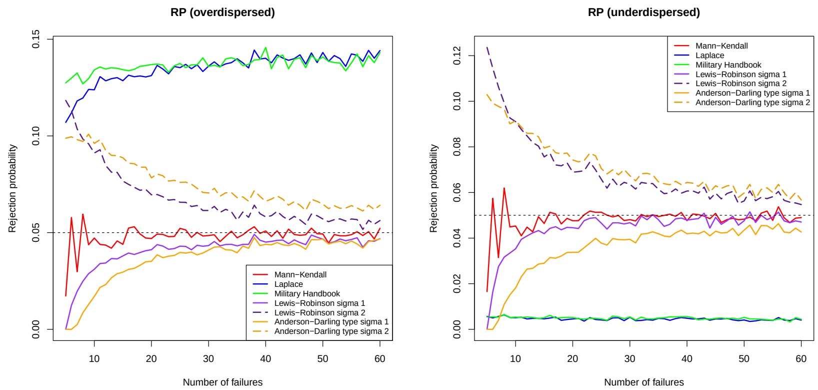


Figure 5.2: Level of the tests simulated as a function of the expected number of events. Data simulated from Weibull RPs with shape parameter 0.75 (overdispersed RP, left plot) and 1.5 (underdispersed RP, right plot), respectively. A nominal level of 0.05 was used.

probability. Conversely, in the right plot they have a way too low rejection probability. We can therefore see that the Military Handbook and Laplace tests are not suitable for a RP situation. This makes sense, since the Military Handbook and Laplace tests use HPP as H_0.

The last plots, Figure 5.3, show how the different tests perform with varying the β parameter for a RP. Firstly, we can look at the Military Handbook and the Laplace tests, which do not perform well for this RP situation. This is something similar to what we saw in Figure 5.2. For $\beta < 1$, the two tests have a rejection probability far too high. Concisely, for $\beta > 1$, the two tests have a way too low rejection probability. We know that for $\beta = 1$, we have an HPP. So, the Military Handbook and Laplace tests are doing great when $\beta = 1$, as we saw in Figure 5.1. The Mann-Kendall test has a slightly low rejection probability for 5 failures but is close to 0.05 for the rest of the plots in Figure 5.3. The only issue with the Mann-Kendall test is that, with very low β values, it has a higher rejection probability than it should. The Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_1$, have a very low rejection probability for 5 failures. With more failures they are performing better. The only thing is that with a low β value, the Lewis-Robinson test with $\hat{\sigma}_1$, does have too low a rejection probability. Concisely, the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_2$, perform too high with 5 and 15 failures. For 30 and 60 failures, they work better. They still have a bit higher rejection probability than they should. In this situation, we would prefer to use $\hat{\sigma}_1$ instead of $\hat{\sigma}_2$, for the Lewis-Robinson and Anderson-Darling tests.

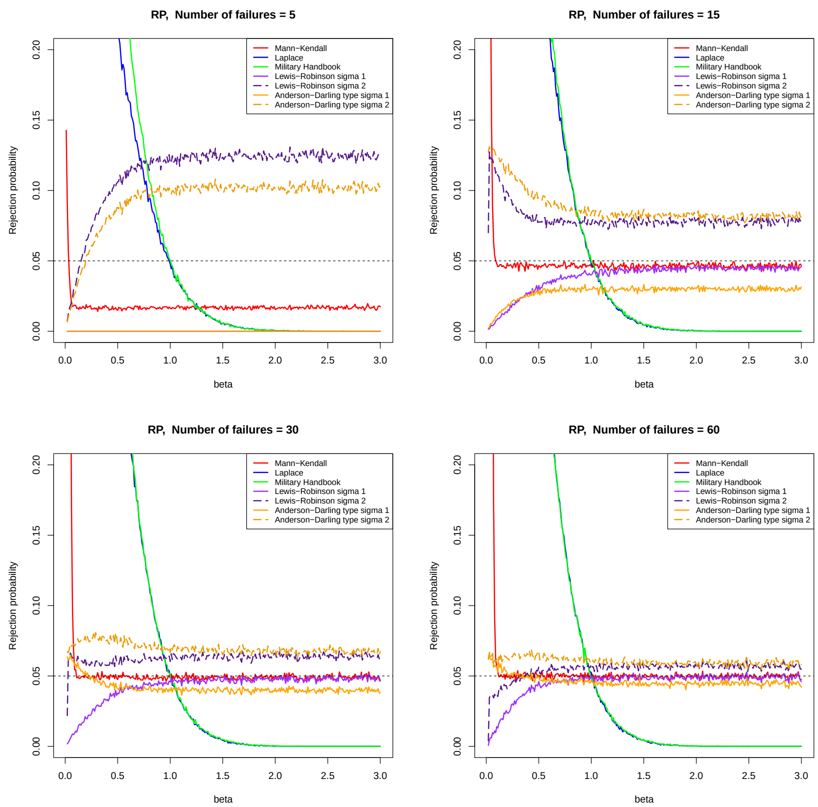


Figure 5.3: Level of the tests simulated as a function of the shape parameter β. Data simulated from Weibull RPs with scale parameter $\alpha = 0.01$ and shape parameter β. A nominal level of 0.05 was used.

5.2 Power Properties

Here we are looking at the power properties, in different situations with trend. Figure 5.4 is simulated with NHPP inversion, the other plots here are simulated with a Weibull TRP model. The intensity function used in these models, is given by $\lambda(t) = abt^{b-1}$. That means in the plots where b is on the x-axis, ideally the rejection probability should be 1 for all b values, other than for $b = 1$. For $b = 1$, the rejection probability should ideally be 0.05. This is true since $b = 1$ means that there is no trend, given the intensity function we have. For $b > 1$, we have an increasing trend. The trend increases more strongly as the b values becomes larger. Conversely, for $b < 1$, we have a decreasing trend. The trend decreases more strongly as the b value becomes smaller.

The first plots for power properties are simulated using NHPP inversion, shown in Figure 5.4.

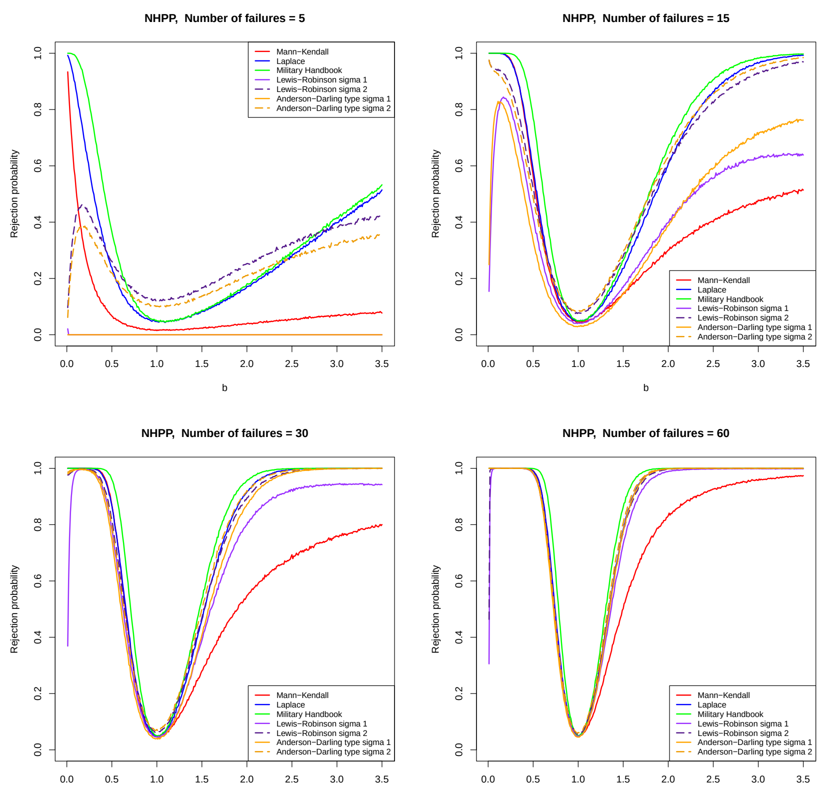


Figure 5.4: Rejection probability of the tests simulated as a function of the trend parameter b. Data simulated from Weibull NHPP with scale parameter 10. A nominal level of 0.05 was used.

In Figure 5.4 we can observe that the Military Handbook and Laplace tests have a rejection probability similar to each other. For 5 failures we can see that they are doing the best of the tests, in this plot. With low b value, they start with rejection probability that are around 1. They go down to around 0.05, before climbing again towards 1. For more failures, they are still the tests to prefer in these situations. That is expected as they use HPP as the H_0. Remember that NHPP is a generalisation of HPP, as illustrated in Figure 2.4.

The Mann-Kendall test does an acceptable job for b values under 1, and better than the other tests for 5 failures. For b values above 1, the Mann-Kendall test has a lower rejection probability than the other tests. The only exception here is for 5 failures, where the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_1$ is lower.

Speaking of the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_1$, when there are 5 failures, they just stay at a rejection probability equal to 0. We saw this in the plots for the level properties, and we will also see it for the TRP plots. Some of this could be explained by the way $\hat{\sigma}_1$ is estimated, compared to $\hat{\sigma}_2$. For 15 failures, they still have a bit too low rejection probability when $b > 1$. For $b < 1$, the tests have a way too low rejection probability, with low values of b. Interestingly with 30 failures, the Anderson-Darling test with $\hat{\sigma}_1$ is doing a bit better than the Lewis-Robinson test with $\hat{\sigma}_1$, when $b > 1$. The Anderson-Darling test with $\hat{\sigma}_1$, also copes better with very low b values, than the Lewis-Robinson test with $\hat{\sigma}_1$. Increasing to 60 failures, the Lewis-Robinson test is about the same with very low b values. What is interesting though, is that also the Lewis-Robinson test with $\hat{\sigma}_2$, does not like very low b values. The reason why it is interesting, is because with fewer failures, the Lewis-Robinson test with $\hat{\sigma}_2$, does not have a problem with that. Even though for 5 failures, the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_2$, is not particular good for very low b values. They are doing better than with $\hat{\sigma}_1$. With $\hat{\sigma}_2$, they have a bit too high rejection probability when $b = 1$. Increasing the failures, the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_2$ are doing about the same thing as the Military Handbook and Laplace tests.

The first TRP plots we are going to look at are shown in Figure 5.5. Recall that overdispersion occurs when $\beta < 1$, indicating more variance than expected in a Poisson process, while underdispersion occurs when $\beta > 1$, indicating less variance.

The first thing to notice in Figure 5.5, is that the Military Handbook and Laplace tests have a higher rejection probability at $b = 1$, than in Figure 5.4. Figure 5.5 shows an overdispersed TRP. As seen for an overdispersed RP in Figure 5.2, the Military Handbook and Laplace tests also have a too high rejection probability here. In the no trend case, recall that for $b = 1$, we have a RP situation. The Military Handbook and Laplace tests perform better than the other test for $b \neq 1$, but since they are higher than they should for $b = 1$, they are not good for an overdispersed TRP situation. This is especially easy to see for 30 and 60 failures.

The Mann-Kendall test has about the same rejection probabilities as in an NHPP situation, as shown in Figure 5.4. However, there are some differences.

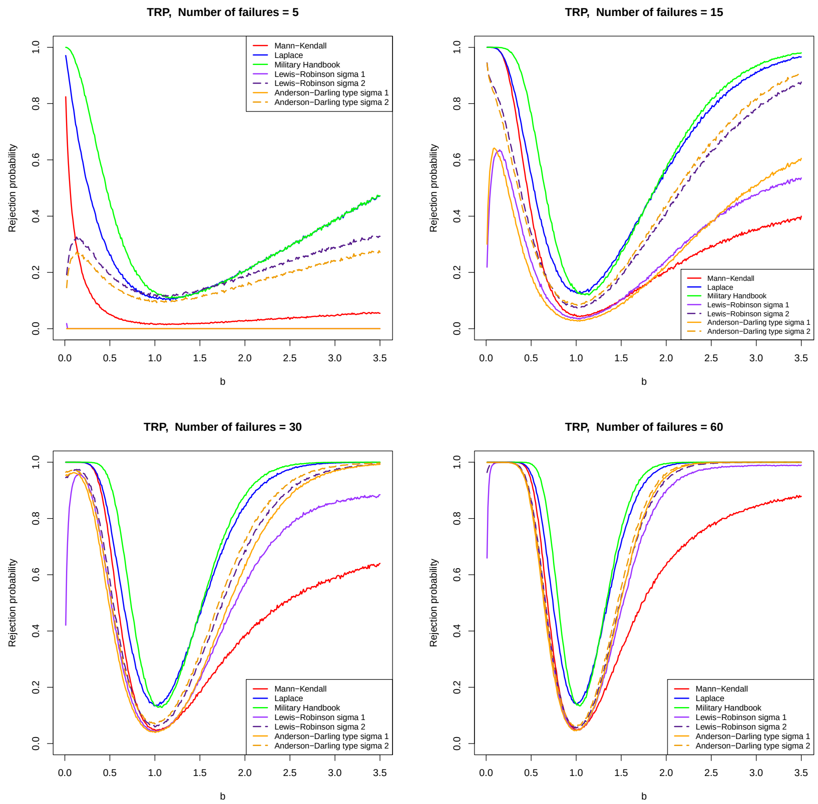


Figure 5.5: Rejection probability of the tests simulated as a function of the trend parameter b. Data simulated from Weibull TRP with shape parameter $\beta = 0.75$ (overdispersed TRP), scale parameters $\alpha = 100$ and $a = 0.1$. A nominal level of 0.05 was used.

The biggest difference is for $b > 1$, the rejection probability is lower than in Figure 5.4, at the same number of failures.

We can see some of the same behaviour of the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_1$, as with the difference seen with the Mann-Kendall test in Figures 5.4 and 5.5. Especially, with 30 failures, the Lewis-Robinson and Anderson-Darling tests have a lower rejection probability than in Figure 5.4, for $b > 1$. In Figure 5.5 with 5 failures, they still have a rejection probability at 0 for all b values.

For the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_2$, they also have a bit lower rejection probability in Figure 5.5 than in Figure 5.4, for $b \neq 1$. They have about the same rejection probability at $b = 1$. These two tests are close to 0.05 at $b = 1$, unlike the Military Handbook and Laplace tests. The

Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_2$ are the ones closest to a rejection probability of 1, for $b \neq 1$, than the other tests. Therefore, in Figure 5.5, the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_2$ seem to be the best tests in this particular situation.

The next plots to look at, is a underdispersed TRP given by Figure 5.6.

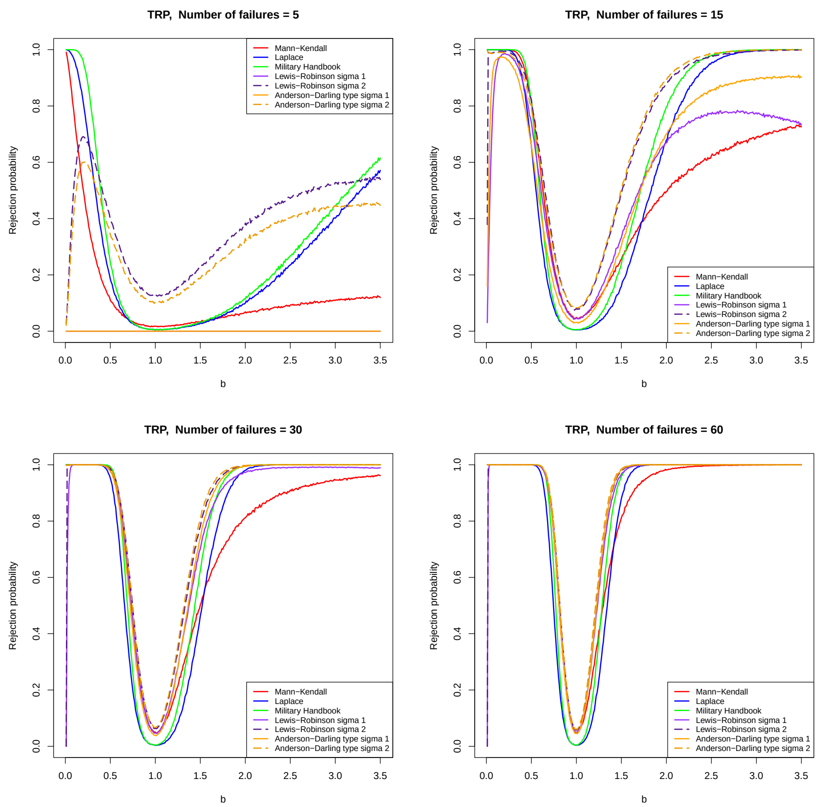


Figure 5.6: Rejection probability of the tests simulated as a function of the trend parameter b. Data simulated from Weibull TRP with shape parameter 1.5 (underdispersed TRP). A nominal level of 0.05 was used.

The first thing to notice in Figure 5.6, is that the Military Handbook and Laplace tests have a much lower rejection probability compared to both Figures 5.4 and 5.5, especially at $b = 1$. Again, we saw that they do not tackle a underdispersed RP in Figure 5.2. Although they still manage to have a high rejection probability at $b > 1.5$ for 30 and 60 failures, they do so for around $b > 2$ for 15 failures.

The behavior of the Mann-Kendall test is still similar in Figure 5.6, compared to the Figures 5.4 and 5.5. However, the Mann-Kendall test is doing a bit better here in Figure 5.6. For 5 failures, at $b > 1$, the test has a bit higher rejection probability than in the other two figures. The rejection probability at $b > 1$ is still too low, though. At $b > 1$, for 15 and more failures, the Mann-Kendall test has a higher rejection probability compared to Figures 5.4 and 5.5. The Mann-Kendall test still manages to have a rejection probability at 0.05, when $b = 1$ and with 15 and more failures.

The Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_1$, still have a rejection probability of 0 for 5 failures. These tests also improve at $b > 1$ for 15 or more failures. The special thing in this figure is that the Lewis-Robinson test with $\hat{\sigma}_1$ starts to get a lower rejection probability at higher b values, from around $b = 2.5$. In addition to that, the Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_1$, have also improved with low values of b, for 15 failures.

For these tests with $\hat{\sigma}_2$, the tests have a higher rejection probability when they should. For 5 failures, the peak at $b < 1$ is higher than in the other two figures. For 5 failures, they also have a higher rejection probability at $b > 1$ than in the other two figures. For 15 and more failures, they are closer to a rejection probability of 1, for $b \neq 1$ than they were in Figure 5.4 for the NHPP situation. At the extremely low b values, the rejection probability is lower than the other two figures, for all the numbers of failures in the plots. This also applies to the Lewis-Robinson test with $\hat{\sigma}_1$.

In conclusion, the tests seem to perform better with an underdispersed TRP than with an NHPP and an overdispersed TRP. This could be due to less variance in the process, which the tests seem to handle more effectively. The exception is obviously the Military Handbook and Laplace tests. Recall that they are made with a Poisson process in mind.

For Figure 5.7 we will look at adjusting the β value, instead of the b value. In Figure 5.7, we have $b = 0.75$. That mean that the rejection probability ideally should be at 1 for every β value in the plot.

The first thing we notice in 5.7, is that none of the tests manages to have a rejection probability close to 1 for all β values in the plots.

To start with, we can look at the Military Handbook and Laplace tests. They generally have a high rejection probability at $\beta < 1$, or at least higher than the other tests. Recall that these two tests, usually have a too high rejection probability in the overdispersed situations. We saw that in Figures 5.2 and 5.5. In contrast, the Military Handbook and Laplace tests have a too low rejection probability with higher β values, especially for $\beta > 1$. For $\beta > 1$, we have an underdispersed situation. Recall that the Military Handbook and Laplace tests have a too low rejection probability in most situations in the underdispersed situation. We saw that in Figures 5.2 and 5.6. Through, the tests manage to get a higher rejection probability at $\beta > 1$ for more failures, especially seen with 60 failures. Although the Military Handbook and Laplace tests manage to have a higher rejection probability than the other tests at $\beta < 1$, remember that they do not retain H_0 often enough when they should in these overdispersed situations.

The Mann-Kendall test is the next test we are going to discuss in Figure 5.7. Firstly what is interesting about the Mann-Kendall test is for 15 or more

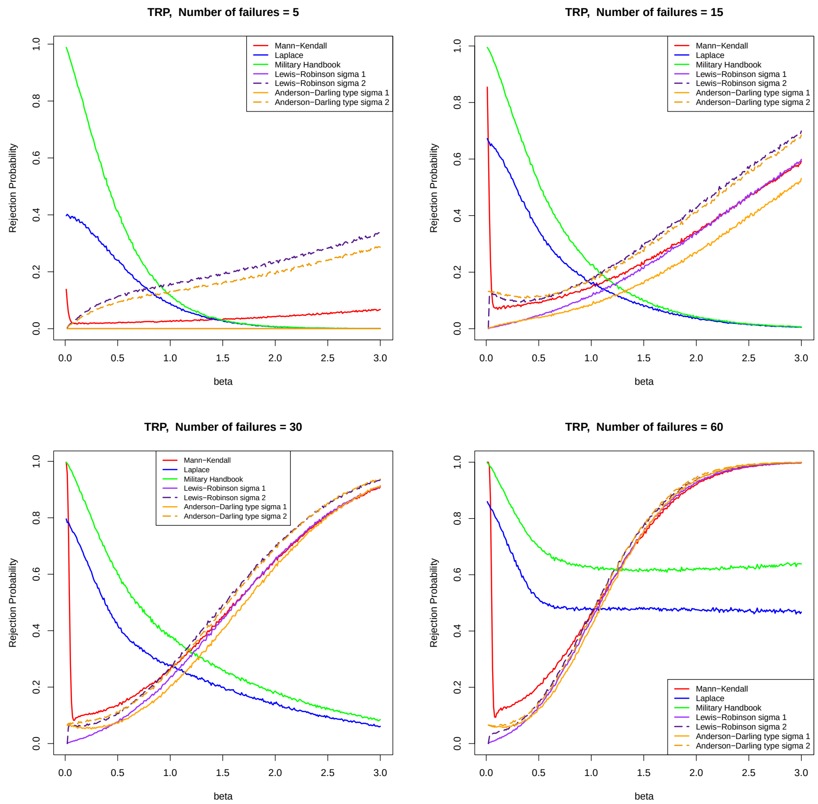


Figure 5.7: Rejection probability of the tests simulated as a function of the trend parameter β. Data simulated from Weibull TRPs with trend parameter 0.75. A nominal level of 0.05 was used.

failures, where the test manages to have a high rejection probability at extremely low β values. However, then the rejection probability goes down immediately. The test manages to increase the rejection probability with higher β values. For 5 failures, the test has a very low rejection probability. However, the rejection probability increases slightly with higher β values in the 5 failures situation. Thus, the test does not archive a high rejection probability for $\beta < 1$, but it performs better with higher β values and for 5 or more failures, especially for $\beta > 1.5$.

The Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_1$ have a similar performance to the Mann-Kendall test. Firstly, these two tests do not handle 5 failures well here either. For 15 or more failures, we see that with higher β values, the tests perform better, much like the Mann-Kendall test. The Laplace test with $\hat{\sigma}_1$ has a higher rejection probability than the Anderson-Darling test with $\hat{\sigma}_1$ for the most part. This is best viewed with 15 failures, where the Laplace test with $\hat{\sigma}_1$ has mostly the same rejection probability for $\beta > 1.5$ as the Mann-Kendall test.

The Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_2$ work better here than with $\hat{\sigma}_1$. The rejection probability is generally higher for the tests with $\hat{\sigma}_2$, but remember that the tests may have a higher rejection probability than they should in some situations. This is best shown in Figure 5.2, where the tests with $\hat{\sigma}_2$ have a too high rejection probability. That is especially shown for low number of failures. However, the tests with $\hat{\sigma}_2$ also have a way too low rejection probability for $\beta < 1$ in Figure 5.7.

From the plots, it may seem that the Military Handbook and Laplace tests generally perform better in Figure 5.7, but we must remember that they generally have a too high rejection probability in the overdispersed situation. So, the other tests may be preferred in this situations, especially for $\beta > 1.5$.

The last plot in this chapter is Figure 5.8, where we will look at a similar situation as in Figure 5.7. The difference is that the trend parameter here is $b = 1.5$, instead of the trend parameter in Figure 5.7, which is $b = 0.75$.

The first thing we see in Figure 5.8 is for 5 failures, we mostly have the same plot as in Figure 5.7. The only noticeable difference is for the Military Handbook test, which has a bit lower rejection probability around $\beta = 0.75$ in Figure 5.8 compared to in Figure 5.7.

The Military Handbook and Laplace tests behave a bit different compared to Figure 5.7. For 15 failures the Military Handbook test drop faster from the very low β values to around $\beta = 0.5$. For $\beta > 0.5$, the test maintains a higher rejection probability than in Figure 5.7. The Laplace test also has a little higher rejection probability at $\beta > 0.5$ in Figure 5.8, than in Figure 5.7. For 30 failures, we see some of the same changes for the Military Handbook and Laplace tests that we did for 15 failures. The Military Handbook test for 30 failures is now increasing the rejection probability with higher β values, from around $\beta = 0.5$. The Laplace test retains approximately the same rejection probability for $\beta > 0.5$, for 30 failures. The test only did that for 60 failures for Figure 5.7. For 60 failures, both the Military Handbook and the Laplace tests are increasing from around $\beta = 0.4$. Both tests are getting close to a rejection probability of 1, from around $\beta > 1$ for the Military Handbook test and from around at $\beta > 1.5$.

The Lewis-Robinson and Anderson-Darling tests with $\hat{\sigma}_1$ behave mostly the same in Figure 5.8 as in Figure 5.7. The noticeable difference between the two figures is that for 15 or more failures, the rejection probability increases faster with increasing β values compared to Figure 5.7. The same is true for the tests using $\hat{\sigma}_2$, including the case with 5 failures.

The Mann-Kendall test performs about the same as in Figure 5.8 as in Figure 5.7. So, while the other tests improve for 15 or more failures, the Mann-Kendall test has about the same rejection probabilities.

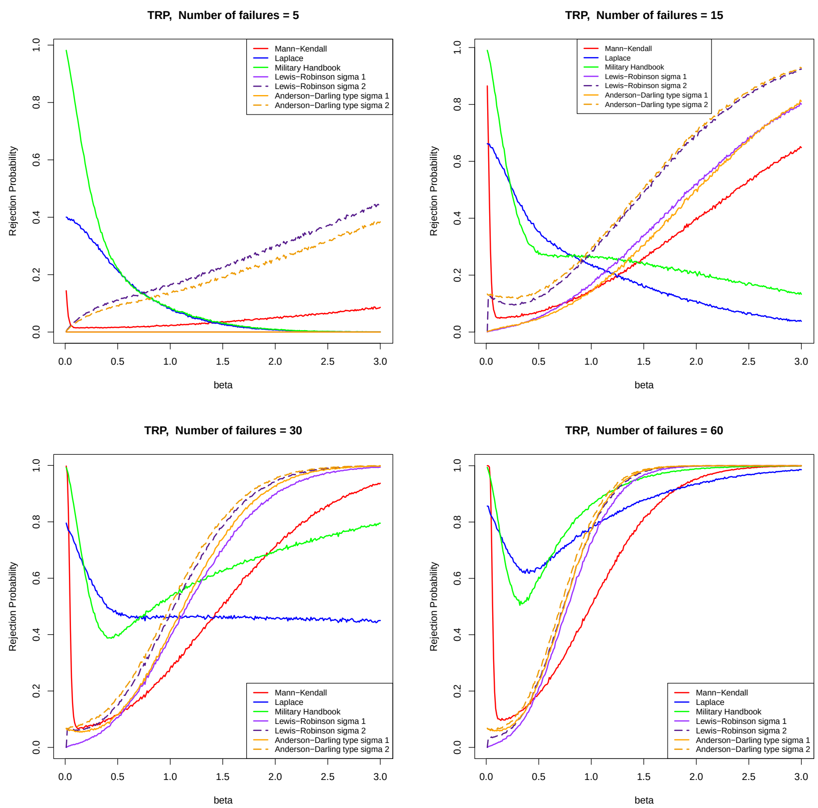


Figure 5.8: Rejection probability of the tests simulated as a function of the trend parameter β. Data simulated from Weibull TRPs with trend parameter 1.5. A nominal level of 0.05 was used.

Chapter 6

Real data examples

In this chapter we are going to look at some real world data. These are the same data that were used in Lindqvist & Kvaløy [14]. The USSH (U.S.S. Halfbeak) data originates from [15, tab. 16.4], and the LHD (Load-Haul-Dump machine) data are from [16, p.341-361]. The first thing we can do with these data is plot them. This gives us a visual representation of the data, as shown in Figure 6.1. A similar analysis was also conducted by [14].

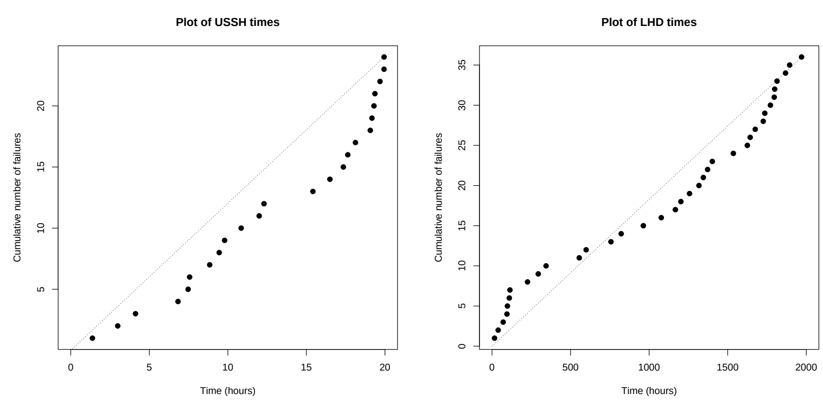


Figure 6.1: Real data plotted. Counts of events as a function of the time in hours. USSH times (left plot) and LHD times (right plot).

The first thing we can see in Figure 6.1 is that there seems to be quite a few points close to 20 hours for the USSH times plot. This could indicate an increasing trend toward the end. For the LHD times, we see some points quite early in the plot, but also some points towards the end of the process. This could indicate a non-monotonic trend, which the Anderson-Darling type test should detect if there is a trend. Recall from Chapter 3.5 that one advantage of

Anderson-Darling type test is that it has strength against both monotonic and non-monotonic trends [6, p.113].

We can then use the tests to calculate whether the tests reject H_0 or not, as shown in Figure 6.2. The data we have is time truncated data, so the Mann-Kendall test may not be the most suitable. However, it is included, so we can see how it performs compared to the other tests.

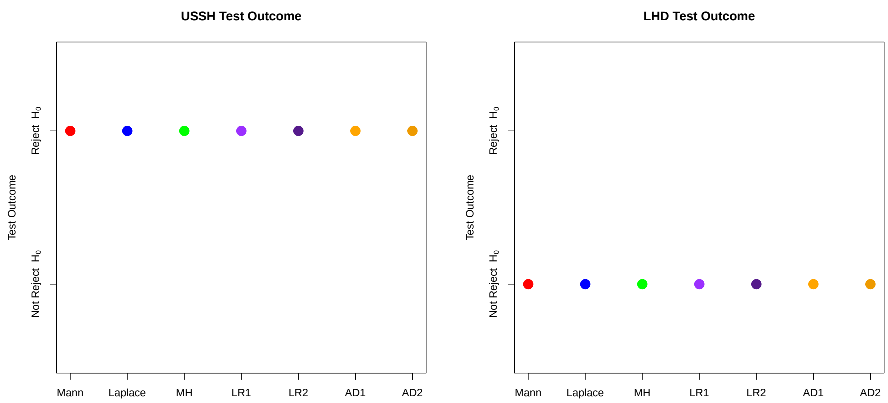


Figure 6.2: Test outcome for USSH (left plot) and for LHD (right plot). The x-axis represents the tests with short names: Mann (Mann-Kendall test), Laplace (Laplace test), MH (Military Handbook test), LR1 (Lewis-Robinson test with $\hat{\sigma}_1$), LR2 (Lewis-Robinson test with $\hat{\sigma}_2$), AD1 (Anderson-Darling type test with $\hat{\sigma}_1$), AD2 (Anderson-Darling type test with $\hat{\sigma}_2$). The y-axis represents whether the test rejects H_0 or not.

In Figure 6.2 we see that every test rejects H_0 for the USSH data. On the other hand, every test does not reject H_0 for the LHD data. Simply put, this means that every test indicates a trend for the USSH data, while every test indicates there is no trend for the LHD data. For the USSH data, the calculation gives us $L_{\text{USSH}} = 2.65$ for the Laplace test. Recall from Chapter 3.1, that a positive L value indicates an increasing trend. This indication aligns with our initial visual interpretation of the data shown in Figure 6.1, where we observed a potential increasing trend. Surprisingly, none of the tests indicated a trend for the LHD data. The Anderson-Darling type test should at least indicate a slight trend. We can check how close the Anderson-Darling type test was to rejecting the H_0 by calculating the p-values. This is shown in Figure 6.3. In this situation, the p-value indicates the strength of evidence against H_0. Since we use a 5% significance level on the test, a p-value lower than 0.05 reject H_0. The lower the p-value, the stronger the evidence against H_0.

From Figure 6.3 we can see the p-value of each test for both the USSH and LHD data. For the p-values from the USSH data, we see that all of the p-

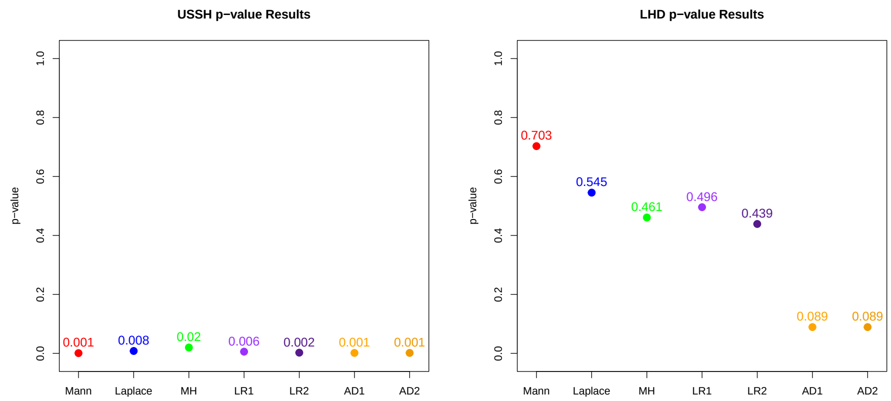


Figure 6.3: The p-values for USSH (left plot) and for LHD (right plot). The x-axis represents the tests with short names: Mann (Mann-Kendall test), Laplace (Laplace test), MH (Military Handbook test), LR1 (Lewis-Robinson test with $\hat{\sigma}_1$), LR2 (Lewis-Robinson test with $\hat{\sigma}_2$), AD1 (Anderson-Darling type test with $\hat{\sigma}_1$), AD2 (Anderson-Darling type test with $\hat{\sigma}_2$). The y-axis represents the p-value outcome from each test.

values are much lower than 0.05. This means that all the tests provide strong evidence against H_0 for the USSH data. For the LHD data, the tests have a much wider range of p-values. All the tests, except for the Anderson-Darling type test, do not provide strong evidence against H_0. The Anderson-Darling type test with both $\hat{\sigma}$, has a p-value of around 0.09. This is not much more than 0.05. Therefore, the Anderson-Darling type test is not far from rejecting H_0. This aligns with our initial visual interpretation of the LHD data in Figure 6.1, with the fact that it could be a non-monotone trend. This is especially notable since the other tests are not as close to rejecting H_0. Recall from Chapter 3.5 that the Anderson-Darling type test also has strength against non-monotonic trends.

Chapter 7

Conclusion

The aim of this thesis was to analyse how different trend tests work in different situations. In particular, the thesis looked at simulated failure truncated data, in different scenarios. There was also some analysis of time truncated data.

Some of the key findings in this thesis are that the Lewis-Robinson and Anderson-Darling type tests perform quite differently with different $\hat{\sigma}$. With the standard $\hat{\sigma}$, referred to in the thesis as $\hat{\sigma}_1$, is not a good $\hat{\sigma}$ with a very low number failures. Another finding is that the Laplace and Military Handbook tests are not good for a Renewal Process (RP) or Trend-Renewal Process (TRP) situation. So unless we are very certain the process is a Poisson situation, the Laplace and Military Handbook tests is not recommended. The Mann-Kendall test, compared to the other tests for the situations studied here, seems to require a higher b value to effectively detect a trend. In this thesis, it was found that all tests studied struggle with high variance, at least for not so pronounced trends, here at $b = 0.75$ and $b = 1.5$. This is shown with low β values in a Weibull TRP situation.

The findings in this thesis provide valuable insights for the application of trend tests in different scenarios. The difference between the choice of different estimates of $\hat{\sigma}$ for the Lewis-Robinson and Anderson-Darling type tests, suggests that careful choice of $\hat{\sigma}$ must be taken especially for low numbers of failures. The observation that the Laplace and Military Handbook tests are not suitable for situations with Renewal Process (RP) or Trend-Renewal Process (TRP), underscores that these tests should only be used where there are a high confidence that the process follows a Poisson process. The observation that the Mann-Kendall test requires a more pronounced increasing trend, which means it requires a higher b value in a Weibull TRP with a intensity function $\lambda(t) = abt^{b-1}$, indicates that this test may not be as sensitive as the other tests at detecting trends. Finally, the fact that all tests struggled with high variance for not so pronounced trends, indicates a potential limitation of these tests in real-world scenarios where there are high variance.

Some limitations of the study is that there is more tests to study. There are also more variants of some of the tests, like the Anderson-Darling type test. Another limitation is that Chapter 5 only simulates failure truncated processes, and in Chapter 6 the real data are only time truncated data. Another limitation is that for the Mann-Kendall test, in the code we assumed that M is approximately normally distributed, even for $n < 10$.

Further research should study why some of the tests change in the rejection probability for very low b and β values. Another thing to examine more closely, is non-monotonic trends. Chapter 5 only simulates failure truncated processes, but it is possible to look at time truncated processes. Further research could analyse more real data to get a better overview of how the different tests react. The data analysed in Chapter 6 is time truncated data, so further research could look at failure truncated data.

For those interested, the R codes for this thesis can be found at Github or go to the url: https://github.com/Benjamin-HN/TrendTestSimulations/. This can also be used for further research.

Bibliography

[1] M. Rausand, A. Barros, and A. Høyland, *System reliability theory: models, statistical methods, and applications*, third edition ed., ser. Wiley series in probability and statistics. Hoboken, NJ: Wiley, 2021.

[2] H. Ascher and H. Feingold, *Repairable systems reliability: modeling, inference, misconceptions and their causes*, ser. Lecture notes in statistics. New York: M. Dekker, 1984, no. v. 7.

[3] M. L. Rizzo, *Statistical computing with R*, second edition ed. Boca Raton: CRC Press, Taylor & Francis Group, 2019.

[4] B. H. Lindqvist, G. Elvebakk, and K. Heggland, "The Trend-Renewal Process for Statistical Analysis of Repairable Systems," *Technometrics*, vol. 45, no. 1, pp. 31–44, Feb. 2003. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1198/004017002188618671

[5] J. T. Kvaløy and B. H. Lindqvist, "TTT-based tests for trend in repairable systems data," *Reliability Engineering & System Safety*, vol. 60, no. 1, pp. 13–28, Apr. 1998. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0951832097000999

[6] ——, "A Class of Tests for Trend in Time Censored Recurrent Event Data," *Technometrics*, vol. 62, no. 1, pp. 101–115, Jan. 2020. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00401706.2019.1605936

[7] P. W. Lewis and D. W. Robinson, "Testing for a Monotone Trend in a Modulated Renewal Process:," Defense Technical Information Center, Fort Belvoir, VA, Tech. Rep., Dec. 1973. [Online]. Available: https://apps.dtic.mil/sti/pdfs/ADA001265.pdf

[8] J. T. Kvaløy and B. Lindqvist, "A class of tests for renewal process versus monontonic and nonmonotonic trend in repairable systems data," in *Mathematical and Statistical Methods in Reliability*, ser. Series on Quality, Reliability and Engineering Statistics, Vol. 7, B. Lindqvist and K. Doksum, Eds. World Scientific Publishing, Singapore, 2003, pp. 401–414.

[9] J. von Neumann, R. H. Kent, H. R. Bellinson, and B. I. Hart, "The Mean Square Successive Difference," *The Annals of Mathematical Statistics*,

vol. 12, no. 2, pp. 153–162, 1941, publisher: Institute of Mathematical Statistics. [Online]. Available: http://www.jstor.org/stable/2235765

[10] H. B. Mann, "Nonparametric Tests Against Trend," *Econometrica*, vol. 13, no. 3, pp. 245–259, 1945, publisher: [Wiley, Econometric Society]. [Online]. Available: https://www.jstor.org/stable/1907187

[11] M. G. Kendall, *Rank correlation methods*, 3rd ed. London: Griffin, 1962.

[12] R. O. Gilbert, *Statistical methods for environmental pollution monitoring*. New York: Van Nostrand Reinhold Co, 1987.

[13] R. E. Walpole, R. H. Myers, J. DeShannon, and K. Ye, *Probability & statistics for engineers & scientists*, ninth edition, global edition ed. Boston Columbus Hoboken: Pearson Education Limited, 2016.

[14] B. H. Lindqvist and J. T. Kvaløy, "New tests for trend in time censored recurrent event data," *Applied Stochastic Models in Business and Industry*, vol. n/a, no. n/a, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/asmb.2848

[15] W. Q. Meeker and L. A. Escobar, *Statistical methods for reliability data*, ser. Wiley series in probability and statistics Applied probability and statistics section. New York Weinheim: Wiley, 1998.

[16] U. Kumar, B. Klefsjö, and S. Granholm, "Reliability investigation for a fleet of load haul dump machines in a Swedish mine," *Reliability Engineering & System Safety*, vol. 26, no. 4, pp. 341–361, Jan. 1989. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0951832089900045