

Tests of homogeneity and change-point inference

A study of sequences of observations with a potential sudden change in their underlying parameters

Elisabeth Nesheim Hagen

Master's Thesis, 1 September 2020



This master's thesis is submitted under the master's programme *Stochastic Modelling, Statistics and Risk Analysis*, with programme option *Statistics*, at the Department of Mathematics, University of Oslo. The scope of the thesis is 60 credits.

The front page depicts a section of the root system of the exceptional Lie group E_8 , projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.

Abstract

We study sequences of independent observations and test whether the observations stem from the same underlying probability distribution. We focus on being able to detect a potential sudden change in the parameters of the distribution, which we call a change-point. Before constructing a test, we define a focus parameter which captures the aspect of the distribution that we want to test. We construct a monitoring process for our focus parameter that converges to a Brownian bridge under the hypothesis of homogeneity. We then use our monitoring process to construct a test statistic for testing homogeneity. We look at the power of our hypothesis test, compared to other tests. We look at how our monitoring process behaves when the null hypothesis is false and suggest a way to estimate the change-point based on the process. We describe two different ways to assess the uncertainty around an estimated change-point with confidence curves.

Acknowledgements

I would like to thank my supervisors Nils Lid Hjort and Céline Cunen. I appreciate the excellent guidance you have given me throughout the writing of this thesis.

I would also like to thank my family and friends for encouraging me.

Contents

Abstract	i
Acknowledgements	iii
Contents	iv
1 Introduction	1
1.1 Related work	2
1.2 Outline of the thesis	3
2 A monitoring process	5
2.1 The monitoring process	5
2.2 The distribution of the monitoring process at a given time-point for the mean	5
2.3 The convergence of the monitoring process for the mean	6
2.4 The convergence of the monitoring process for a general focus parameter	8
2.5 Special cases	11
3 Defining a test statistic	17
3.1 The hypothesis	17
3.2 A test statistic	17
3.3 Application: Battle-deaths in interstate conflict	20
4 Power	23
5 The behaviour of the monitoring process	27
5.1 Parametric vs non-parametric tests	27
5.2 When the hypothesis of homogeneity is true	27
5.3 When the hypothesis of homogeneity is false	28
6 Estimating the change-point	31
7 Confidence sets	35
7.1 Confidence sets using profiled log-likelihood and deviance	35
7.2 Constructing confidence sets using our test statistic M_n	36
8 Application: British coal-mining disasters	41
8.1 An analysis assuming normally distributed observations	42
8.2 Assuming a Poisson distribution	44
8.3 Assuming observations from a Gamma-Poisson distribution	46
9 Application: Tirant lo Blanch	49
10 Supplementary remarks	53
A Some R code	55
Bibliography	59

Chapter 1

Introduction

Our starting point is that we have a sequence of independent observations and want to test whether all the observations stem from the same distribution. For a scientist studying a sequence of data-points, the data are often observations in time, and we will think of the observations as observations in time. We assume that the observations come from the same parametric family $f(y, \theta)$. For example, we might assume that our observations are normally distributed, but the underlying parameters of the distribution might be different for each observation. We denote the vector of the parameters of the distribution in question θ . For a sequence of observations $y_1, \dots, y_n$, let θ_i correspond to observation y_i . Our null hypothesis is that θ is homogeneous, meaning that all θ_i are equal

$$H_0 : \theta_1 = \dots = \theta_n.$$

We define a focus parameter denoted by T , which is the parameter we want to test, for example the mean or the standard deviation. To test the homogeneity of the focus parameter we need to define a test statistic. Before we define a test statistic we construct a monitoring process, denoted H_n , that converges to a Brownian Bridge process under the hypothesis of homogeneity. Our monitoring process will give us information about whether the estimated focus parameter appears significantly non-homogeneous. We will use the limit process of H_n under the hypothesis of homogeneity to construct a test statistic with a known distribution. For an index τ of the sequence of observations, the process $H_n(\tau)$ at τ indicates whether there is a significant difference in the estimated focus parameter based on the observations before and after τ . We reject the hypothesis of homogeneity if the monitoring process takes large values. To illustrate the monitoring process H_n under homogeneity we simulate a sequence of 100 standard normal values and plot the corresponding H_n process in Figure [1.1](#)

A change-point (also called break-point or regime shift) is where a sequence of observations has a sudden change in the parameters of its underlying distribution.

For cases where the null hypothesis fail, our focus will be on detecting the failure of the null hypothesis when there is one change-point such that $y_1, \dots, y_\tau \sim f(y, \theta_L)$ and $y_{\tau+1}, \dots, y_n \sim f(y, \theta_R)$, where θ_L and θ_R are two different vectors of parameters. For instance, there could be a difference in the mean before and after the change-point, a jump in the standard deviation or there could be some other more subtle difference.

When we estimate a change-point we also want to measure the uncertainty. To measure uncertainty we describe two different ways to construct confidence curves. From a confidence curve we can read off the confidence set for any given confidence level α .

For the calculations done in R, I have written my own code, and not used ready-made packages.

The theory of change-points has many applications, and in many fields of study researchers can use this theory to study specific structural breaks in their observations. A researcher might, depending on the situation, assume that there is an underlying common distribution from which their observations has been generated, which is done in many applications of statistics, but even when this assumption is justified, it may be worth considering the uncertainty of this underlying constancy.

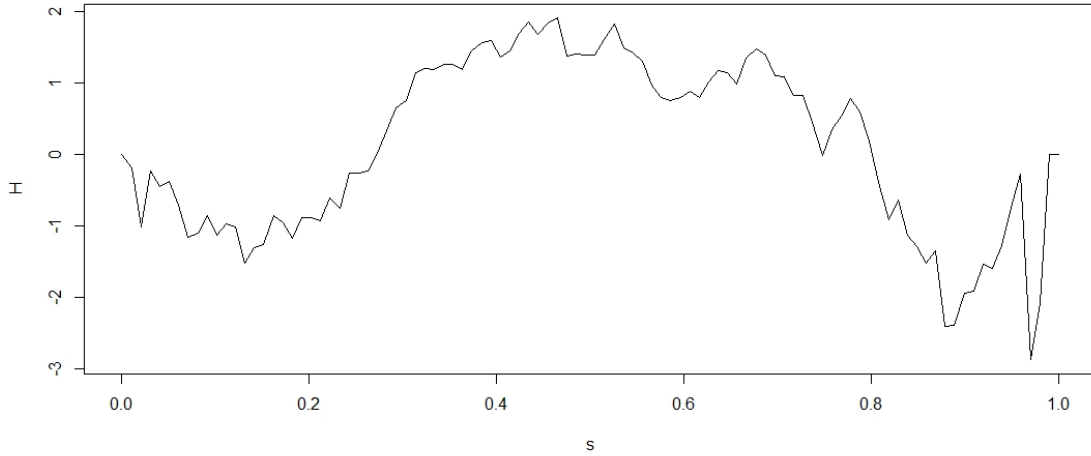


Figure 1.1: The monitoring process $H_n^*(s, \mu)$ for a sequence of $n = 100$ data-points drawn from the standard normal distribution. The focus parameter is the mean. The process converges to a Brownian motion for an increasing number of observations.

1.1 Related work

The article by [Cunen et al., 2018](#) presents two general methods for estimating a break-point along with measures of uncertainty, assuming that the observations $y_1, \dots, y_\tau$ stem from a model $f(y, \theta_L)$ and the observations $y_{\tau+1}, \dots, y_n$ stem from $f(y, \theta_R)$. The first method is to construct a goodness-of-fit test, test the homogeneity of observations before and after all potential change-points separately, and form a confidence curve. This is what we will do as well, using the hypothesis test that we will construct. The second method is to use profile log-likelihood to estimate the change-point, and then form the deviance function from the profile to find the confidence curve, we look at this method in more detail in Chapter [7](#). The article also finds confidence curves for the degree of change at the change-point, which we will not do in our analysis. In the context of the first confidence curve method the authors investigate the use of monitoring processes. A monitoring process that converges to a Brownian bridge is constructed based on log-likelihood, which is different from how we construct our monitoring process. The methods are demonstrated in four applications, we will analyse two of the same cases. The first application is the number of British coal-mining disasters from 1851 to 1962. They assume the observations are Poisson distributed and has one change-point. Using profile log-likelihood they find that the maximum likelihood estimate of the change-point and its confidence curve. We will also analyse the British coal-mining disasters. We will apply our test of homogeneity with three different assumptions about the distribution, and demonstrate that homogeneity is very unlikely. We find two likely change-points and corresponding confidence curves. The second application in the article is the 487 chapter long novel *Tirant lo Blanch* from the 1460s. The aim is to pinpoint where the second author (Marti Joan de Galba) took over for the first author (Joanot Martorell). We will test the homogeneity of the text, and use our test statistic to guess where the change is.

The article by [Cunen et al., 2020](#) investigates the number of battle deaths in 95 interstate wars, from 1823 to 2003. There has been various research that has supported the claim that the world has become more peaceful over time. Other researchers argue that war fatalities have remained stable over time. [Cunen et al., 2020](#) argues that war fatalities has decreased. We will test the homogeneity of the mean and the standard deviation of the battle deaths without assuming that the observations stem from a specific distribution, and will not find reason to reject the null hypothesis of homogeneity.

The article by [Koning and Hjort, 2002](#), uses the score function to construct a monitoring process for testing the homogeneity of a vector of parameters, which converges to a vector of Brownian bridges under the null hypothesis. The article also considers weighted monitoring processes.

1.2 Outline of the thesis

In Chapter 2 we only consider theory under the null hypothesis, and assume that we have a sequence of observations that are independent and identically distributed. We construct the monitoring process denoted $H_n(\tau, T)$. We use Donsker's theorem to find the limit process of $H_n(\tau, T)$. We find the correct variance involved in the monitoring process using influence functions. We show that the process converges to a standard normal distribution for a given index τ , and prove that the process as a whole converges to a Brownian motion process. We end the chapter with some special cases of the monitoring process.

In Chapter 3 we define the test statistic M_n as the maximum absolute value of the monitoring process H_n . We let the statistic M be the maximum absolute value of the Brownian Bridge process that H_n converges to. We show that M_n converges to M . The distribution of M_n depends on the number of observations n , especially for small n , but since M_n converges to M , we can find the distribution of M , and compare M_n to the quantiles of M . We find the distribution of M and the distribution of M_n for a few given values of n by simulation. We name the hypothesis test using the statistic M_n the M-test, and end the chapter by applying the M-test to data of Battle-deaths in interstate conflict, where we test whether the likelihood of violent battles has changed over time.

In Chapter 4 we simulate the power of the M-test and compare it to the likelihood ratio test and a Chi-squared test.

In Chapter 5 we compare how our monitoring process behaves depending on whether null hypothesis is true or false for different focus parameters. We look at different versions of the process by letting the focus parameter be the variance $T = \sigma^2$, the standard deviation $T = \sigma$ and the logarithm of the standard deviation $T = \log(\sigma)$ for simulated data-sets. For a homogeneous sequence of values, we notice that the process is almost identical regardless of the choice of T . On the other hand, for a data-set where the variance has a change-point, the processes is less similar for different focus parameters.

In Chapter 6 we look at how we can estimate a change-point using our monitoring process, and suggest using the index of the most extreme value of the process as an estimate.

In Chapter 7 we describe a method of finding confidence curves using profile log-likelihood and a method using the M-test. We compare the realized coverage level of confidence sets constructed using the M-method with the theoretical confidence level. We notice that the coverage level of M_n for small n are smaller than what we expect when n is large.

In Chapter 8 we use our theory to analyse British coal-mining disasters. We look at the number of disasters per year, first by not assuming any specific distribution, secondly by assuming a Poisson distribution and thirdly by assuming a Gamma-Poisson distribution. There is strong evidence in favour of rejecting the null hypothesis of homogeneity in all three cases. The likelihood of disaster does not appear to be constant over time. We find two likely change-points.

In Chapter 8 we analyse the novel *Tirant lo Blanch*. The novel has two authors and we try to identify in which chapter the second author started writing based on the length of the words used.

Chapter 2

A monitoring process

In this chapter we are going to define a monitoring process that we will later use in hypothesis testing of homogeneity. For the purpose of describing the process under the null hypothesis, we assume that we have independent identically distributed observations. For a sequence of observations $Y = \{y_i\}_{i=1}^n$, let y_i be drawn from a parametric family $f(y, \theta)$, where θ has dimension p . In later chapters we will not assume that the observations are identically distributed, but rather that y_i come from a parametric family $f(y, \theta_i)$, where θ_i is the underlying parameter of the observation y_i .

2.1 The monitoring process

Say we have a focus parameter T , where $T = T(F)$ is functional mapping from a distribution to a scalar. Let $\hat{T} = T(F_n)$ be the estimate of T , where F_n is the empirical distribution. For an index τ indicating an observation y_τ , let $\hat{T}_L$ be the estimate of T from $y_1, \dots, y_\tau$, and let $\hat{T}_R$ be the estimate of T from $y_{\tau+1}, \dots, y_n$. We think of $\hat{T}_L$ and $\hat{T}_R$ as the estimates of T to the left and to the right of τ respectively. We will study the process

$$H_n(\tau, T) = \frac{\hat{T}_L - \hat{T}_R}{\{\hat{\kappa}_L^2/\tau + \hat{\kappa}_R^2/(n - \tau)\}^{1/2}}. \quad (2.1)$$

For parts of the analysis, instead of using the index τ , it will make more sense to define the process such that it starts at 0 and ends at 1. Let $s = \tau/n$, the process is then defined as

$$H_n^*(s, T) = \frac{\hat{T}_L - \hat{T}_R}{\{\hat{\kappa}_L^2/ns + \hat{\kappa}_R^2/n(1 - s)\}^{1/2}}. \quad (2.2)$$

The value $\hat{\kappa}^2$ is the estimated variance involved, which we will define later in this chapter. We use the notation $H_n(T)$ and $H_n^*(T)$ when we consider s (or τ) a fixed constant. We will show that for a given s , $H_n^*(T)$ converges to a standard normal random variable for increasing n . We will also show that the whole process $H_n^*(s, T)$ converges to a Brownian motion process.

Before reaching the formal definition of the process and general proofs, we first look at the special case where T is the expected value of the observations. We prove the convergence of $H_n^*(T)$ in Section 2.2 and the convergence of the process $H_n^*(s, T)$ in Section 2.3 for the special case $T = E(y)$. In Section 2.4 we prove the convergence of the process $H_n^*(s, T)$ for a general functional T . In Section 2.5 we describe some relevant special cases of the general theory.

2.2 The distribution of the monitoring process at a given time-point for the mean

We let τ be a known constant, we evaluate $H_n(T)$ for the special case when T is the mean, $T = \mu$. In this section we derive the limit distribution of $H_n(T)$ for this case by using the Central Limit Theorem. The general result is that $H_n(T)$ converges to a standard normal distribution for most functionals T we might define.

Theorem 2.1 (Central Limit Theorem). Casella and Berger, 2002 Let $Y_1, Y_2, \dots$ be a sequence of independent identically distributed random variables with $E Y_i = \mu$ and $0 < \text{Var } Y_i = \sigma^2 < \infty$. Define $\bar{Y}_n = (1/n) \sum_{i=1}^n Y_i$. Let $G_n(y)$ denote the cumulative distribution function of $\sqrt{n}(\bar{Y}_n - \mu)/\sigma$. Then, for any y , $-\infty < y < \infty$,

$$\lim_{n \rightarrow \infty} G_n(y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx,$$

that is, $\sqrt{n}(\bar{Y}_n - \mu)/\sigma$ has a limiting standard normal distribution.

Consider a sequence of independent random variables y_i from a family of distributions $f(y, \theta)$, with expected value μ and finite variance σ^2 .

Result 2.2. Suppose we observe an independent identically distributed sample $y_1, \dots, y_n$, with mean μ and finite variance σ^2 , then

$$H_n^*(\mu) = \frac{\hat{\mu}_L - \hat{\mu}_R}{\{\hat{\sigma}_L^2/sn + \hat{\sigma}_R^2/n(1-s)\}^{1/2}}$$

tends to the standard normal distribution when n increases.

Proof. The numerator of H_n^* is

$$\hat{\mu}_L - \hat{\mu}_R = \frac{1}{ns} \sum_{i \leq ns} y_i - \frac{1}{n(1-s)} \sum_{i > ns} y_i,$$

and

$$E[\hat{\mu}_L - \hat{\mu}_R] = 0.$$

The variance of $\hat{\mu}_L$ and $\hat{\mu}_R$ are

$$\text{Var}(\hat{\mu}_L) = \frac{\sigma^2}{ns}, \quad \text{and} \quad \text{Var}(\hat{\mu}_R) = \frac{\sigma^2}{n(1-s)}$$

We then have

$$\text{Var}(\hat{\mu}_L - \hat{\mu}_R) = \frac{\sigma^2}{ns} + \frac{\sigma^2}{n(1-s)},$$

For a consistent estimator $\hat{\sigma}^2$ of σ^2 , say the the maximum likelihood estimator, $\hat{\sigma}^2$ converges in probability to σ^2 . From the Central limit theorem, the sample average converges to a normal distribution, and we have that

$$H_n^*(\mu) = \frac{\hat{\mu}_L - \hat{\mu}_R}{\{\hat{\sigma}_L^2/sn + \hat{\sigma}_R^2/n(1-s)\}^{1/2}}$$

tends to the standard normal distribution when n increases, using $\hat{\kappa}^2 = \hat{\sigma}^2$. □

2.3 The convergence of the monitoring process for the mean

We now consider the whole monitoring process $H_n(s, T)$. In this section we prove that the process converges to a Brownian motion process for the special case when the focus parameter T is the expected value, as before $\hat{\kappa}^2$ is the estimated variance of the focus parameter.

Definition 2.3 (Brownian motion). Ross, 1983 A stochastic process $[W(s), s \geq 0]$ is said to be a standard *Brownian motion* if

1. $W(0) = 0$
2. $\{W(s), s \geq 0\}$ has stationary independent increments

3. for every $s > 0$, $W(s)$ is normally distributed with mean 0 and variance s .

Theorem 2.4 (Donsker). *Let $Y_1, Y_2, \dots$ be a sequence of independent and identically distributed random variables, with mean 0 and variance 1. The empirical partial-sum process*

$$A_n(s) = \frac{1}{\sqrt{n}} \sum_{i \leq \lfloor ns \rfloor} Y_i \quad \text{for } s \in [0, 1]$$

converges in distribution to a standard Brownian motion process $W = \{W(s) : s \in [0, 1]\}$

Here $\lfloor ns \rfloor$ is the floor of ns , and is an integer. The convergence takes place in the space $D[0, 1]$ of all right-continuous functions $y : [0, 1] \rightarrow \mathbb{R}$ with left hand limits, with the Skorokhod topology.

A Brownian bridge is a (standardised) Brownian motion process that takes the value zero at both the start and the end. We denote a Brownian bridge by $W^0(s)$.

Result 2.5. *Suppose we observe an independent identically distributed sample $y_1, \dots, y_n$, with mean μ and finite variance σ^2 , then $H_n(s, \mu)$ converges to a Brownian motion process in distribution*

$$H_n^*(s, \mu) \rightarrow_d \frac{W^0(s)}{\sqrt{s(1-s)}}.$$

Proof. Assume independent identically distributed observations from a distribution with mean μ and variance σ^2 . We have our test statistic

$$H_n^*(s, \mu) = \frac{\hat{\mu}_L - \hat{\mu}_R}{\{\hat{\sigma}_L^2/sn + \hat{\sigma}_R^2/n(1-s)\}^{1/2}}.$$

The numerator can be written

$$\begin{aligned} \hat{\mu}_L - \hat{\mu}_R &= \frac{1}{ns} \sum_{i \leq ns} (\mu + z_i) - \frac{1}{n(1-s)} \sum_{i > ns} (\mu + z_i) \\ &= \frac{1}{ns} \sum_{i \leq ns} z_i - \frac{1}{n(1-s)} \sum_{i > ns} z_i \end{aligned}$$

where the expectation of z_i is 0. Hence, without loss of generality we can assume $\mu = 0$. We know that $\hat{\sigma}_L$ and $\hat{\sigma}_R$ converge to the same value σ . Writing the process with sums, we have

$$H_n^*(s, \mu) = \frac{\frac{1}{ns} \sum_{i \leq ns} z_i}{\frac{\hat{\sigma}_L}{\sqrt{n}} \{1/s + 1/(1-s)\}^{1/2}} - \frac{\frac{1}{n(1-s)} (\sum_{i \leq n} z_i - \sum_{i \leq ns} z_i)}{\frac{\hat{\sigma}_R}{\sqrt{n}} \{1/s + 1/(1-s)\}^{1/2}}.$$

Where we have an expression containing only the observations before the change-point minus an expression containing only observations after the change-point. Multiply and divide with σ , and with some rearranging, we write

$$H_n^*(s, \mu) = \frac{\frac{\sigma}{s\hat{\sigma}_L\sqrt{n}} \sum_{i \leq ns} z_i/\sigma}{\{1/s + 1/(1-s)\}^{1/2}} - \frac{\frac{\sigma}{(1-s)\hat{\sigma}_R\sqrt{n}} (\sum_{i \leq n} z_i/\sigma - \sum_{i \leq ns} z_i/\sigma)}{\{1/s + 1/(1-s)\}^{1/2}}.$$

Writing this using $A_n(s) = \frac{1}{\sqrt{n}} \sum_{i \leq ns} z_i/\sigma$, we have

$$\begin{aligned} H_n^*(s, \mu) &= \frac{\frac{\sigma}{\hat{\sigma}_L} A_n(s)}{s\{1/s + 1/(1-s)\}^{1/2}} - \frac{\frac{\sigma}{\hat{\sigma}_R} (A_n(1) - A_n(s))}{(1-s)\{1/s + 1/(1-s)\}^{1/2}} \\ &= \frac{\frac{\sigma}{\hat{\sigma}_L} A_n(s) \sqrt{s(1-s)}}{s} - \frac{\frac{\sigma}{\hat{\sigma}_R} (A_n(1) - A_n(s)) \sqrt{s(1-s)}}{(1-s)} \\ &= \frac{\frac{\sigma}{\hat{\sigma}_L} A_n(s) \sqrt{(1-s)}}{\sqrt{s}} - \frac{\frac{\sigma}{\hat{\sigma}_R} (A_n(1) - A_n(s)) \sqrt{s}}{\sqrt{(1-s)}}. \end{aligned}$$

Since $\hat{\sigma}$ is a consistent estimator of σ [Casella and Berger, 2002], both $\sigma/\hat{\sigma}_L$ and $\sigma/\hat{\sigma}_R$ converges to 1.

Using Donsker's theorem, $A_n(s)$ converges to a standard Brownian motion $W(s)$. From Slutsky's theorem $H_n^*(s, \mu)$ converges in distribution, for increasing n ,

$$H_n^*(s, \mu) \rightarrow_d \frac{\sqrt{(1-s)W(s)}}{\sqrt{s}} - \frac{\sqrt{s}(W(1) - W(s))}{\sqrt{(1-s)}}. \quad (2.3)$$

We can write the expression as a single fraction,

$$H_n^*(s, \mu) \rightarrow_d \frac{W(s) - sW(1)}{\sqrt{s(1-s)}} = \frac{W^0(s)}{\sqrt{s(1-s)}}. \quad (2.4)$$

□

2.4 The convergence of the monitoring process for a general focus parameter

In this section we prove the general version of Result 2.5. In order to show that the process $H_n(\tau, T)$ converges to a Brownian bridge process for a more general functional T , we use influence functions. We also use influence functions to find the correct $\hat{\kappa}^2$ involved. We find the influence function of the mean before we show how a general functional T can be expressed with the influence function. We first define the influence function.

Definition 2.6. [Huber, 1981] For a cdf F and $\epsilon > 0$, the *influence function* of a statistic $T = T(F)$ (a functional of the distribution) at a point y is

$$IF(y, F, T) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [T((1-\epsilon)F + \epsilon\delta_y) - T(F)] = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [T(F_\epsilon) - T(F)],$$

where $Y \sim F_\epsilon$ if

$$Y \sim \begin{cases} F & \text{with probability } 1 - \epsilon \\ y & \text{with probability } \epsilon \end{cases}$$

and where $F_\epsilon = (1-\epsilon)F + \epsilon\delta_y$, that is, F_ϵ is a mixture of F and a point mass in y . We can think of F_ϵ as a contaminated version of F .

The influence function can be seen as the Gâteaux derivative of the functional T , that shows how a small contamination of the distribution in question is affecting T .

We first find the influence function of the mean. For some distribution denoted F , let the statistic $\mu(F)$ be the mean, the influence function is

$$IF(y, F, \mu(F)) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [\mu(F_\epsilon) - \mu(F)].$$

Since expectation is a linear map, we can write $\mu(F_\epsilon)$ as

$$\mu(F_\epsilon) = \mu((1-\epsilon)F + \epsilon y) = (1-\epsilon)\mu(F) + \epsilon y.$$

We then have

$$\begin{aligned} IF(y, F, \mu(F)) &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [(1-\epsilon)\mu(F) + \epsilon y - \mu(F)] \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [-\epsilon\mu(F) + \epsilon y] \\ &= y - \mu(F). \end{aligned}$$

A general focus parameter expressed with the influence function

Let $T(F_n)$ be the estimate of the test statistic $T(F)$, estimated from the empirical distribution. Assume that $T(F_n)$ is consistent, meaning that

$$T(F_n) \rightarrow T(F)$$

in probability.

We use the same arguments as [Huber, 1981](#) to show the convergence of $T(F_n)$. Let $Y_1, \dots, Y_n$ be iid random variables. If T is Gâteaux differentiable at F and sufficiently regular, it can be linearized near F in terms of $IF(y, F, T)$. The leading terms of the Taylor expansion of $T(F_n)$ are

$$T(F_n) = T(F) + \int IF(y, F, T)[F_n(dy) - F(dy)] + \dots,$$

since

$$\int IF(y, F, T)F(dy) = 0$$

we have

$$\begin{aligned} \sqrt{n}(T(F_n) - T(F)) &= \sqrt{n} \int IF(y, F, T)F_n(dy) + \epsilon_n \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n IF(y_i, F, T) + \epsilon_n. \end{aligned}$$

Where ϵ_n denote the remaining terms which are asymptotically negligible and converge in probability to zero. By the central limit theorem, the leading term on the right-hand side is asymptotically normal.

We use the notation $o_p(1)$ for a sequence of random variables that converge to zero in probability. A sequence X_n converges in probability to zero or is bounded in probability at "rate" R_n if

$$X_n = o_p(R_n)$$

which means that

$$X_n = R_n o_p(1).$$

Lemma 2.7. *For an appropriately differentiable functional T ,*

$$\sqrt{n}(T(F_n) - T(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n IF(y_i, F, T) + o_p(1).$$

is asymptotically normal with mean 0 and variance κ^2 ([Vaart, 1998](#)). The general estimate of the statistic T is

$$\hat{T} = T + \frac{1}{n} \sum_{i=1}^n IF(y_i, T) + o_p(1).$$

We can use lemma [2.7](#) to formally define the monitoring process

$$H_n^*(s, T) = \frac{\hat{T}_L - \hat{T}_R}{\{\hat{\kappa}_L^2/sn + \hat{\kappa}_R^2/n(1-s)\}^{1/2}}.$$

In general, for a sequence of observations $y_1, \dots, y_m$ we define $\hat{\kappa}^2$ as

$$\hat{\kappa}^2 = \text{Var} \left(\frac{1}{\sqrt{m}} \sum_{i=1}^m IF(y_i, F, \hat{T}) \right) = \text{Var} \left(IF(y, F, \hat{T}) \right) = \text{E} \left(IF^2(y, F, \hat{T}) \right).$$

For the sequence of observations $y_1, \dots, y_\tau$, with $ns = \tau$ observations

$$\hat{\kappa}_L^2 = \text{Var} \left(\frac{1}{\sqrt{ns}} \sum_{i=1}^{ns} IF(y_i, F, \hat{T}) \right),$$

and similarly, for the sequence of observations $y_{\tau+1}, \dots, y_n$, with $n(1-s) = n - \tau$ observations

$$\hat{\kappa}_R^2 = \text{Var} \left(\frac{1}{\sqrt{n(1-s)}} \sum_{i=1}^{n(1-s)} IF(y_i, F, \hat{T}) \right).$$

Definition 2.8. For a set of observations $y_1, \dots, y_n$ and a functional T of the distribution,

$$H_n(\tau, T) = \frac{\hat{T}_L - \hat{T}_R}{\{\hat{\kappa}_L^2/\tau + \hat{\kappa}_R^2/(n - \tau)\}^{1/2}}. \quad (2.5)$$

Where $\hat{T} = T(F_n)$ denote the estimate of the statistic T . The value $\hat{\kappa}^2$ is the variance of the influence function of $\hat{T}$,

$$\hat{\kappa}^2 = \text{Var} \left(IF(y, F, \hat{T}) \right).$$

If we use the true value T in place of the estimate $\hat{T}$, we have

$$\kappa^2 = \text{Var} (IF(y, F, T)).$$

Given an index τ of the observations, the estimates $\hat{T}$ and $\hat{\kappa}^2$ based on $y_1, \dots, y_\tau$ are denoted $\hat{T}_L$ and $\hat{\kappa}_L^2$ respectively. Similarly, $\hat{T}_R$ and $\hat{\kappa}_R^2$ are the estimates based on $y_{\tau+1}, \dots, y_n$. Let $c \leq \tau/n \leq d$ for a suitable interval $[c, d]$ inside $(0, 1)$. Define $s = \tau/n$, the index τ divided by the length of the sequence of observations n . Writing the expression as a function of s , we have

$$H_n^*(s, T) = \frac{\hat{T}_L - \hat{T}_R}{\{\hat{\kappa}_L^2/sn + \hat{\kappa}_R^2/n(1-s)\}^{1/2}}, \quad (2.6)$$

with $c \leq s \leq d$.

In Section 2.3 we proved that the process $H_n^*(s, T)$ converges to a Brownian motion process for the special case when T is the mean. We now prove this result for a general functional T .

Theorem 2.9. *Given a sequence of independent identically distributed observations and a general functional T satisfying lemma 2.7 for all s , where $0 < s < 1$,*

$$H_n^*(s, T) \rightarrow_d \frac{W^0(s)}{\sqrt{s(1-s)}}. \quad (2.7)$$

Proof. When lemma 2.7 holds the nominator

$$\hat{T}_L - \hat{T}_R = T + \frac{1}{ns} \sum_{i \leq ns} IF(y_i, F, T) + \epsilon_\tau - \left(T + \frac{1}{n(1-s)} \sum_{i > ns} IF(y_i, F, T) + \epsilon_{n-\tau} \right)$$

has an asymptotically normal distribution with expected value 0. We write

$$H_n^*(s, T) = \frac{\frac{1}{ns} \sum_{i \leq ns} IF(y_i, F, \hat{T}) - \frac{1}{n(1-s)} \sum_{i > ns} IF(y_i, F, \hat{T})}{\{\hat{\kappa}_L^2/sn + \hat{\kappa}_R^2/n(1-s)\}^{1/2}}.$$

With similar calculations as we have done before for Result 2.5, we can write this as

$$H_n^*(s, T) = \frac{\frac{\kappa}{s\hat{\kappa}_L\sqrt{n}} \sum_{i \leq ns} IF(y_i, F, \hat{T})/\kappa}{\{1/s + 1/(1-s)\}^{1/2}} - \frac{\frac{\kappa}{(1-s)\hat{\kappa}_R\sqrt{n}} (\sum_{i \leq n} IF(y_i, F, \hat{T})/\kappa - \sum_{i \leq ns} IF(y_i, F, \hat{T})/\kappa)}{\{1/s + 1/(1-s)\}^{1/2}}.$$

From lemma 2.7 $A_n(s) = \frac{1}{\sqrt{n}} \sum_{i \leq ns} IF(y_i, F, \hat{T})/\kappa$ has expected value 0 and variance 1, and we write

$$H_n^*(s) = \frac{\frac{\kappa}{\hat{\kappa}_L} A_n(s)}{s\{1/s + 1/(1-s)\}^{1/2}} - \frac{\frac{\kappa}{\hat{\kappa}_R} (A_n(1) - A_n(s))}{(1-s)\{1/s + 1/(1-s)\}^{1/2}}.$$

The estimate $\hat{\kappa}$ converges to κ , and hence $\kappa/\hat{\kappa}_L$ and $\kappa/\hat{\kappa}_R$ converges to one. We apply Donsker's theorem to $A_n(s)$, and we have that the monitoring process converges in distribution to a Brownian bridge for a general focus parameter T ,

$$H_n^*(s, T) \rightarrow_d \frac{\sqrt{(1-s)}W(s)}{\sqrt{s}} - \frac{\sqrt{s}(W(1) - W(s))}{\sqrt{(1-s)}} = \frac{W^0(s)}{\sqrt{s(1-s)}}. \quad (2.8)$$

□

2.5 Special cases

In this section we now show some more special cases of $H_n^*(s, T)$.

The monitoring process for the mean

In Section 2.2 (Result 2.2) we showed that the monitoring process for the mean

$$H_n^*(\mu) = \frac{\hat{\mu}_L - \hat{\mu}_R}{\{\hat{\kappa}_L^2/sn + \hat{\kappa}_R^2/n(1-s)\}^{1/2}}$$

tends to the standard normal distribution when $\hat{\kappa}^2 = \hat{\sigma}^2$. In Section 2.4 we showed that the influence function of the mean is

$$IF(y, F, \mu(F)) = y - \mu(F).$$

We can now state from more general theory that

$$\hat{\kappa}^2 = \text{Var}[y_i - \hat{\mu}(F)] = \text{E}[(y_i - \hat{\mu}(F))^2],$$

and we get the same estimate as before, $\hat{\kappa}^2 = \hat{\sigma}^2$.

For independent and identically distributed observations $y_1, \dots, y_i, \dots, y_n$, let $\hat{\mu}_n = \frac{1}{n} \sum y_i$ be the estimated mean. We start with the influence function of the expected value, and find an expression for the estimated mean by writing

$$\begin{aligned} IF(y_i, F, \mu(F)) &= y_i - \mu(F) \\ y_i &= \mu(F) + IF(y_i, F, \mu) \\ \frac{1}{n} \sum_{i=1}^n y_i &= \mu(F) + \frac{1}{n} \sum_{i=1}^n IF(y_i, F, \mu) \\ \hat{\mu}_n &= \mu(F) + \frac{1}{n} \sum_{i=1}^n IF(y_i, F, \mu). \end{aligned}$$

We can now write

$$\hat{\mu}_n - \mu(F) = \frac{1}{n} \sum_{i=1}^n IF(y_i, F, \mu(F))$$

and

$$\sqrt{n}(\hat{\mu}_n - \mu(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n IF(y_i, F, \mu(F)).$$

On the left-hand side we have an expression that we know converges by the central limit theorem.

The monitoring process for the variance

Let the focus parameter T be the variance, $T(F) = \sigma^2$. For observations $y_1, \dots, y_n$, let $\bar{y} = 1/n \sum_i y_i$, and let

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

be the estimator of the variance σ^2 . We want to find the expression of $\hat{\kappa}^2$ of the process

$$H_n^*(s, \sigma^2) = \frac{\hat{\sigma}_L^2 - \hat{\sigma}_R^2}{\{\hat{\kappa}_L^2/sn + \hat{\kappa}_R^2/n(1-s)\}^{1/2}}.$$

Result 2.10. Suppose we observe an independent normally distributed sample $y_1, \dots, y_n$, with mean μ and finite variance σ^2 , then

$$H_n^*(\sigma^2) = \frac{\hat{\sigma}_L^2 - \hat{\sigma}_R^2}{\left\{ \frac{2\hat{\sigma}_L^4}{ns} + \frac{2\hat{\sigma}_R^4}{n(1-s)} \right\}^{1/2}}$$

tends to the standard normal distribution for a given s for increasing n .

Proof. We use the fact that the expression $(n-1)\hat{\sigma}_n^2/\sigma^2$ has a chi squared distribution with $n-1$ degrees of freedom [Casella and Berger, 2002]. Let X_M be a chi squared distributed random variable with $m = n-1$ degrees of freedom. Then, we can write $\hat{\sigma}_n^2$ as

$$\hat{\sigma}_n^2 = \frac{\sigma^2 X_M}{n-1}.$$

Let X_1 be a random variable following a Chi-squared distribution χ_1^2 , let $x_1, \dots, x_m$ be a sample drawn from χ_1^2 . We know the expected value and variance of x_i ,

$$\begin{aligned} E(x_i) &= 1 \\ \text{Var}(x_i) &= 2. \end{aligned}$$

From central limit theorem, we have that

$$\sqrt{m} \left(\frac{\sum_{i=1}^m x_i}{m} - 1 \right)$$

converges in distribution as m increases to $N(0, 2)$. The sum $\sum_{i=1}^m x_i$ has a χ^2 -distribution with m degrees of freedom, let's name it x_M . Since m/n converges to 1, we also have

$$\sqrt{n} \left(\frac{x_M}{m} - 1 \right) \rightarrow_d N(0, 2).$$

Multiply the expression by σ^2 , and we have

$$\sqrt{n} \left(\frac{\sigma^2 x_M}{m} - \sigma^2 \right) = \sqrt{n} (\hat{\sigma}_n^2 - \sigma^2) \rightarrow_d N(0, 2\sigma^4).$$

Hence, $H_n^*(s, \sigma^2)$ converges to the standard normal distribution when $\hat{\kappa}^2 = 2\hat{\sigma}^4$. □

We find a more general expression of the variance $\hat{\kappa}^2$ of the estimated variance by finding the influence function of the estimated variance.

Result 2.11. *Suppose we observe an independent distributed sample $y_1, \dots, y_n$, with mean μ and finite variance σ^2 , then*

$$H_n^*(\sigma^2) = \frac{\hat{\sigma}_L^2 - \hat{\sigma}_R^2}{\left\{ \frac{E[(\hat{\mu} - y)^4] - \hat{\sigma}_L^4}{ns} + \frac{E[(\hat{\mu} - y)^4] - \hat{\sigma}_R^4}{n(1-s)} \right\}^{1/2}}$$

tends to the standard normal distribution for a given s for increasing n .

Proof. The influence function is defined

$$IF(y, F, \sigma^2) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [\sigma^2(F_\epsilon) - \sigma^2(F)].$$

Expectation is a linear map, hence we have that

$$\begin{aligned} E_{F_\epsilon}(Y^2) &= (1 - \epsilon) E_F(Y^2) + \epsilon y^2 \\ [E_{F_\epsilon}(Y)]^2 &= [(1 - \epsilon) E_F(Y) + \epsilon y]^2 \\ &= (1 - \epsilon)^2 E_F(Y)^2 + 2\epsilon y(1 - \epsilon) E_F(Y) + \epsilon^2 y^2. \end{aligned}$$

We use this to find the variance of the contaminated distribution,

$$\begin{aligned} \sigma^2(F_\epsilon) &= E_{F_\epsilon}(Y^2) - [E_{F_\epsilon}(Y)]^2 \\ &= (1 - \epsilon)\sigma^2(F) + \epsilon(1 - \epsilon) E_F(Y)^2 - 2\epsilon(1 - \epsilon)y E_F(Y) + \epsilon(1 - \epsilon)y^2. \end{aligned}$$

We then get

$$\sigma^2(F_\epsilon) - \sigma^2(F) = -\epsilon\sigma^2(F) + \epsilon(1 - \epsilon) E_F(Y)^2 - 2\epsilon(1 - \epsilon)y E_F(Y) + \epsilon(1 - \epsilon)y^2.$$

The influence function is

$$\begin{aligned} IF(y, F, \sigma^2) &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [-\epsilon\sigma^2(F) + \epsilon(1 - \epsilon) E_F(Y)^2 - 2\epsilon(1 - \epsilon)y E_F(Y) + \epsilon(1 - \epsilon)y^2] \\ &= \lim_{\epsilon \rightarrow 0} [-\sigma^2(F) + (1 - \epsilon) E_F(Y)^2 - 2(1 - \epsilon)y E_F(Y) + (1 - \epsilon)y^2] \\ &= E_F(Y)^2 - 2y E_F(Y) + y^2 - \sigma^2(F) \\ &= (E_F(Y) - y)^2 - \sigma^2(F). \end{aligned}$$

From our theory, writing μ instead of $E_F(Y)$, we then have

$$\begin{aligned} \kappa^2 &= \text{Var} [IF(y, F, \sigma^2)] = \text{Var} [(\mu - y)^2 - \sigma^2(F)] = \text{Var}[(\mu - y)^2] \\ &= E[(\mu - y)^4] - E[(\mu - y)^2]^2 = E[(\mu - y)^4] - \sigma^4. \end{aligned}$$

□

What the expectation $E[(\mu - y)^4]$ is depends on which assumptions we work with. If y_i is assumed to be standard normal, we find the variance using the Chi-squared distribution,

$$\kappa^2 = \text{Var}[y^2] = 2$$

We can use moments to find

$$E[(\mu - y)^4] = E[Y^4] - 4\mu E[Y^3] + 6\mu^2 E[Y^2] - 3\mu^4.$$

Under the assumption of normality

$$\kappa^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4.$$

Say we have identically distributed observations $y_i, \dots, y_i, \dots, y_n$, with unknown parameters. We can express the mean as

$$\mu = \hat{\mu}_n - \epsilon_n,$$

where ϵ_n is the small difference between $\hat{\mu}_n$ and μ . we have

$$\begin{aligned} IF(y_i, F, \sigma^2) &= (y_i - \hat{\mu} + \epsilon_n)^2 - \sigma^2 \\ &= (y_i - \hat{\mu})^2 + 2\epsilon_n(y_i - \hat{\mu}) + \epsilon_n^2 - \sigma^2. \end{aligned}$$

We can now find an expression of the estimated variance, let $\epsilon_n^* = 2\epsilon_n(y_i - \hat{\mu}) + \epsilon_n^2$, we have

$$\begin{aligned} (y_i - \hat{\mu})^2 &= \sigma^2 + IF(y_i, F, \sigma^2) - \epsilon_n^* \\ \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2 &= \sigma^2 + \frac{1}{n} \sum_{i=1}^n IF(y_i, F, \sigma^2) - \bar{\epsilon}_n^* \\ \hat{\sigma}_n^2 &= \sigma^2 + \frac{1}{n} \sum_{i=1}^n IF(y_i, F, \sigma^2) - \bar{\epsilon}_n^*. \end{aligned}$$

Since $\hat{\mu}$ converges to μ in probability for increasing n , $\bar{\epsilon}_n^*$ is sufficiently small for large n . We have

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \approx \frac{1}{\sqrt{n}} \sum_{i=1}^n IF(y_i, F, \sigma^2).$$

The monitoring process for the standard deviation

When the exact asymptotic variance is not straightforward to find, we can find its approximation more easily using the Delta method. The delta method is an easy way to use the weak convergence of a sequence of random variables to find the weak convergence of a transformation of the random variables.

Definition 2.12 (Delta method). [Casella and Berger, 2002] Let Y_n be a sequence of random variables that satisfy $\sqrt{n}(Y_n - \theta) \rightarrow N(0, \sigma^2)$ in distribution. For a given function g and a specific value of θ , suppose that $g'(\theta)$ exist and is not 0. Then

$$\sqrt{n}[g(Y_n) - g(\theta)] \rightarrow N(0, \sigma^2[g'(\theta)]^2).$$

We assume normally distributed observations and let the the statistic T be the standard deviation $T(F) = \sigma$. Since we know that

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \rightarrow_d N(0, 2\sigma^4)$$

we can use the delta method to find the variance of the estimated standard deviation $\hat{\sigma}$, Let $g(\sigma^2) = \sqrt{\sigma^2}$, then $g'(\sigma^2) = \frac{1}{2\sigma}$, and we have

$$[g'(\sigma^2)]^2 = \frac{1}{4\sigma^2}.$$

Hence, by the Delta method

$$\sqrt{n}(\hat{\sigma}_n - \sigma) \rightarrow_d N(0, 2\sigma^4 \frac{1}{4\sigma^2}) = N(0, \sigma^2/2).$$

We then achieve Result 2.13 by using

$$\hat{\kappa}^2 = \frac{\hat{\sigma}^2}{2}.$$

Result 2.13. Suppose we observe an independent normally distributed sample $y_1, \dots, y_n$, with mean μ and finite standard deviation σ , then

$$H_n(\sigma) = \frac{\hat{\sigma}_L - \hat{\sigma}_R}{\left\{ \frac{\hat{\sigma}_L^2}{2ns} + \frac{\hat{\sigma}_R^2}{2n(1-s)} \right\}^{1/2}}$$

tends to the standard normal distribution for a given s when n increases.

Instead of approximating the asymptotic variance of the estimated standard deviation, we can also find it exactly by using what we know about the distribution of the estimated variance, we know that the estimated variance is

$$\hat{\sigma}^2 = \frac{\sigma^2 X_m}{n-1},$$

where $m = n-1$ and X_m has a chi-squared distribution. To find the variance of the estimated standard deviation $\hat{\sigma}$, we write

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \frac{\sigma V_m}{\sqrt{n-1}},$$

where V_m has a chi distribution with m degrees of freedom, hence, we can find the exact variance of $\hat{\sigma}$ by finding

$$\text{Var } \hat{\sigma}_n = \frac{\sigma^2}{(n-1)} \text{Var}(V_m) = \frac{\sigma^2}{(n-1)} \left(n-1 - \frac{2 \Gamma((n-1)/2)^2}{\Gamma(n/2)^2} \right),$$

and then use this to find the limit distribution of

$$\sqrt{n}(\hat{\sigma} - \sigma).$$

The monitoring process for the logarithm of the standard deviation

We can also find the asymptotic variance of a function of the standard deviation $g(\hat{\sigma}) = \log \hat{\sigma}$.

Result 2.14. Suppose we observe an independent normally distributed sample $y_1, \dots, y_n$, with mean μ and finite standard deviation σ , then

$$H_n(\log \sigma) = \frac{\log \hat{\sigma}_L - \log \hat{\sigma}_R}{\left\{ \frac{1}{2ns} + \frac{1}{2n(1-s)} \right\}^{1/2}}$$

tends to the standard normal distribution for a given s when n increases.

Proof. We use that

$$\sqrt{n}(\hat{\sigma} - \sigma) \rightarrow_d N(0, \sigma^2/2)$$

we have

$$g'(\sigma) = \frac{\partial}{\partial \sigma} \log(\sigma) = \frac{1}{\sigma}$$

$$[g'(\sigma)]^2 = \frac{1}{\sigma^2}$$

hence, by the Delta Method,

$$\sqrt{n}(g(\hat{\sigma}) - g(\sigma)) \rightarrow_d N\left(0, \frac{\sigma^2}{2} \frac{1}{\sigma^2}\right) = N\left(0, \frac{1}{2}\right).$$

We then achieve Result [2.14](#) by using

$$\kappa^2 = \frac{1}{2}$$

which does not depend on the standard deviation σ . □

The influence function of the probability of an observation being larger than a given value

Let $Y_1, \dots, Y_n$ be a sequence of independent identically distributed random variables with expectation μ and variance σ^2 . Let the statistic T be the probability that we observe a value larger than some given value y_0 , $T = \Pr(Y > y_0)$. Let $I(y > y_0)$ be the indicator function that is 0 when $y < y_0$ and 1 when $y > y_0$. We then have that $T(F_\epsilon)$ is

$$\begin{aligned} \Pr(Y > y_0|F_\epsilon) &= \begin{cases} \Pr(Y > y_0|F) & \text{with probability } (1 - \epsilon) \\ I(y > y_0) & \text{with probability } \epsilon \end{cases} \\ &= (1 - \epsilon) \Pr(Y > y_0|F) + \epsilon I(y > y_0) \end{aligned}$$

The influence function is

$$\begin{aligned} IF(y, F, \Pr(y_0)) &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [\Pr(Y > y_0|F_\epsilon) - \Pr(Y > y_0|F)] \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [(1 - \epsilon) \Pr(Y > y_0|F) + \epsilon I(y > y_0) - \Pr(Y > y_0|F)] \\ &= [I(y > y_0) - \Pr(Y > y_0|F)] \end{aligned}$$

Chapter 3

Defining a test statistic

We want to test the homogeneity of sequences of observations. We test the homogeneity of a focus parameter by using the theory from Chapter 2.

3.1 The hypothesis

For a sequence of observations $y_1, \dots, y_n$, we assume that observation i is drawn from a distributions $f(y, \theta_i)$. We want to test if θ_i is the same for all indices i . Our null hypothesis H_0 is that $\theta_1 = \dots = \theta_n$. Denote the set of distributions satisfying the null hypothesis $\mathcal{F}_0$. We want to test H_0 against some well defined alternative H_A . The alternative hypothesis is that the sample $y_1, \dots, y_n$ are drawn from a parametric family $f(y, \theta_i)$, where it is not true that $\theta_1 = \dots = \theta_n$, which means that there are at least two indices, i and j , such that $\theta_i \neq \theta_j$. Denote the alternative set of distribution not satisfying the null hypothesis $\mathcal{F}_A$. Our focus will be on being able to detect the failure of H_0 when there are one (and only one) change-point, happening at an index τ , such that $y_1, \dots, y_\tau$ stem from $f(y, \theta_L)$, and $y_{\tau+1}, \dots, y_n$ from $f(y, \theta_R)$, which is only a small part of $\mathcal{F}_A$. The more difficult failures of H_0 to detect might be cases where only a single θ_i is different, creating just one outlier, or cases where θ oscillates between some values.

We will test the homogeneity of θ by testing the homogeneity of a focus parameter T , of the distribution. We will assume that all our observations stem from the same distribution $f(y, \theta)$ if we accept that our focus parameter T is homogeneous for all observations. If for example we test whether the mean $T(F) = \mu$ is constant, and there is a true non-constancy in the standard deviation, we might conclude that we have homogeneity, when we in fact do not. In order to detect the failure of H_0 we need to test the feature of the distribution where the failure happens. We might test the feature we care most about, or we can test all the parameters of the distribution separately. If we for example have normally distributed observations, we should consider both the mean and the variance in some way, maybe by evaluating two different foci T .

3.2 A test statistic

We now decide to use the most extreme value of the process $H_n^*(s, T)$ as a test statistic. To make sure that $\hat{T}_L$ and $\hat{T}_R$ are not estimated based on very few observations, we are interested in the the most extreme value of $H_n^*(s, T)$ for s in a given interval. We decide to work with s in the interval $0.1 < s < 0.9$, using the notation in definition [2.8](#) we let $c = 0.1$ and $d = 0.9$.

Definition 3.1. Given a sequence of observations $y_1, \dots, y_n$, we define a test statistic M_n , using $H_n^*(s, T)$ defined in Chapter 2 (definition [2.8](#)), as

$$M_n = \max_{0.1 \leq s \leq 0.9} |H_n^*(s, T)| = \max_{0.1 \leq s \leq 0.9} \left| \frac{\hat{T}_L - \hat{T}_R}{\{\hat{\kappa}_L^2/ns + \hat{\kappa}_R^2/n(1-s)\}^{1/2}} \right|.$$

We know from Chapter 2 that $H_n^*(s, T)$ converges to a Brownian motion process under the null hypothesis. We will use the the limiting distribution of the $H_n^*(s, T)$ process to derive the limit distribution of M_n under the hypothesis of homogeneity.

Definition 3.2. Let $|W^0(s)|$ be the absolute value of $W^0(s)$, where $W^0(s)$ is a Brownian Bridge process, define

$$q(s) = \frac{|W^0(s)|}{\sqrt{s(1-s)}}, \quad (3.1)$$

$$M = \max_{0.1 \leq s \leq 0.9} q(s). \quad (3.2)$$

Result 3.3. When the null hypothesis of homogeneity is true, our test statistic M_n converges in distribution to M , the maximum of the the absolute value of a Brownian Bridge process,

$$M_n \rightarrow_d M. \quad (3.3)$$

Proof. We know that

$$H_n^*(s) \rightarrow_d \frac{W^0(s)}{\sqrt{s(1-s)}},$$

since $g(H_n) = \max |H_n|$ is a continuous function ([Billingsley, 1968](#))

$$M_n \rightarrow_d M = \max_{c_0 \leq s \leq d_0} \frac{|W^0(s)|}{\sqrt{s(1-s)}}.$$

□

In later chapters we will refer to the hypothesis test where we use the statistic $M_n(T)$ as the M-test. A large M_n value indicates that the null hypothesis is false, since expected value of the Brownian motion process is zero. We need the distribution of the statistic M in order to be able to compare observed M_n values to quantiles of M .

We want to find the distribution of M empirically. Drawing values from M means drawing Brownian motion processes, and it is not possible to simulate a continuous-time process, we can only simulate discrete-time processes. We therefore simulate processes on K points, $0 < s_1 < \dots < s_k < \dots < 1$, in stead of the interval $[0, 1]$. We do this by first drawing a sequence of values $z_1, \dots, z_k, \dots, z_K$ from the standard normal distribution, here we use $K = 10^4$. Every step length $s_{k+1} - s_k$ of our approximate Brownian motion is $1/K$. We get a value m from the approximated distribution of M by by calculating

$$m = \max_{0.1 \leq s_k \leq 0.9} \frac{|A_K(s_k) - s_k A_K(1)|}{s_k(1-s_k)}$$

of a sequence $z_1, \dots, z_K$ where A_K is an approximate Brownian Motion, similarly as in Chapter 2

$$A_K(s_k) = \frac{1}{\sqrt{K}} \sum_{k \leq \lfloor ns_k \rfloor} Z_k \quad \text{for } s_k \in [0, 1].$$

After drawing 10^5 values m , we have a good estimate of the distribution of the statistic M and its percentiles. We have plotted the distribution of M in [Figure 3.2](#) and listed some percentiles in [Table 3.2](#).

In applications we will test the constancy of finite sequences, we therefore want some intuition of how fast M_n converges to M to make a good evaluation of an observed M_n . We briefly investigate how fast the statistic M_n converges to M by simulation. We do this by simulating a large number of M_n values for the specific case when $y_1, \dots, y_n \sim N(0, 1)$ and $T = \mu$. For $n = 100, 200, 500$, we create 10^4 simulations of M_n , and compare the estimated quantiles of M_n to qualities of M . From the results in [Table 3.1](#), we notice that M_n is larger when we have fewer observations, which can lead us to reject the null hypothesis more often than we intend.

We also look at how the expected value of M_n depends on n , especially for small values of n . We simulate the 10^4 values of M_n for a given n and find the mean value of M_n . Each M_n is found from a sequence of n simulated standard normal values. We do this for different small values of n . We plot the mean of M_n given n in [Figure 3.1](#).

Percentile	M_{100}	M_{200}	M_{500}	M
0.99	3.992	3.749	3.606	3.574
0.95	3.218	3.097	3.012	3.037
0.90	2.857	2.801	2.730	2.758
0.80	2.480	2.462	2.424	2.450

Table 3.1: We simulated the distribution of M_{100} , M_{200} and M_{500} . The table shows the percentile values for the M_n statistic when the sequence of observations are from the standard normal distribution, with $T = \mu$ and $s = 0.5$.

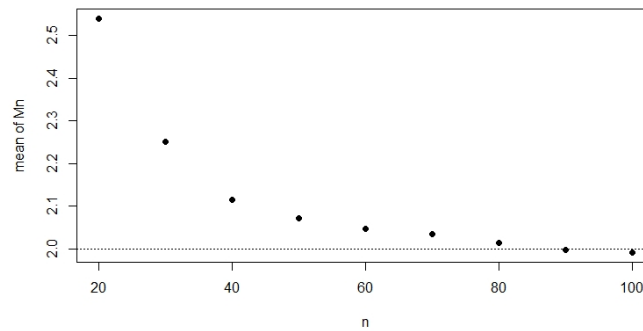


Figure 3.1: The expected value of the statistic M_n given n .

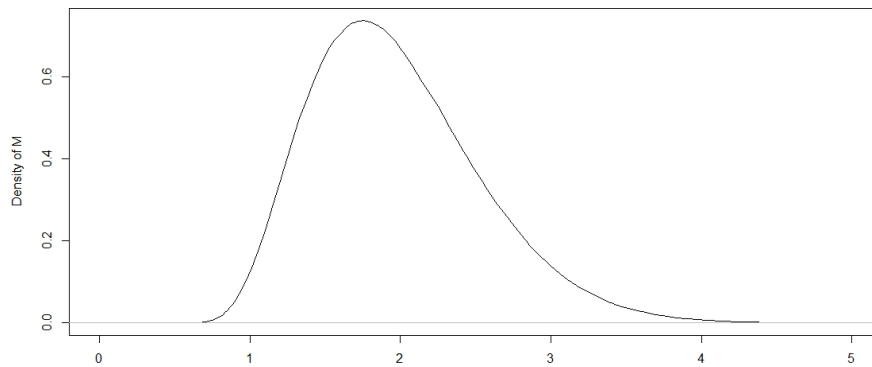


Figure 3.2: Simulated distribution of the statistic M , which M_n converges to.

Percentile	0.01	0.05	0.10	0.20	0.50	0.80	0.90	0.95	0.99
Value	1.012	1.210	1.334	1.513	1.930	2.450	2.758	3.037	3.574

Table 3.2: Percentile values of distribution of M (approximated).

Other test statistics

The statistic M_n that we have chosen to work with, is of course one of many alternative test statistics we could have chosen. Instead, one could for example be interested in the area under the curve of the process $H_n(s, T)$. The integral is a continuous function, and therefore we know what this statistic converges to. One could use the integral of the process squared

$$\int_{0.1}^{0.9} (H_n^*(s, T))^2 ds \rightarrow_d \int_{0.1}^{0.9} \frac{|W^0(s)|^2}{s(1-s)} ds.$$

Or one might be interested in the integral of the absolute value of the process $H_n(s, T)$.

$$\int_{0.1}^{0.9} |H_n^*(s, T)| ds \rightarrow_d \int_{0.1}^{0.9} \frac{|W^0(s)|}{\sqrt{s(1-s)}} ds.$$

3.3 Application: Battle-deaths in interstate conflict

We look at the number of people killed in 95 interstate wars between 1823 and 2003. The data is provided by The Correlates of War Project. The numbers range from 1000 killed (the minimum by definition) to 16,634,907 (the recorded battle deaths of World War Two) ([Sarkees and Wayman, 2010](#)).

We wonder if the likelihood of violent battles has changed over time. Since there has been relatively few large interstate wars after 1945, some have argued that the world has become more peaceful. We want to test whether the number of people killed in battle per war follow the same trend over time. The analysis of Clauset ([Clauset, 2018](#)) indicates that the pattern of conflict have been stationary over time. His analysis indicates that the postwar pattern of peace would need to endure at least another 100 to 140 years to become a statistically significant trend. A trend that would lead to rejecting homogeneity. On the other hand, [Cunen et al., 2020](#) find a likely change-point for the most severe wars, with more than 7061 deaths, to be 1965. for the full data-set they find a likely change-point in 1950.

Wars are vastly different in size, and the largest wars are many orders of magnitude larger than a “typical” war. The number of battle deaths is known to follow a right-skewed distribution with a heavy tail, we will therefore analyse $y_i = \log(z_i)$ the logarithm of the number of people killed. We consider the number of deaths in battle alone, ignore other characteristics of conflicts, and make no assumption about the distribution of the observations. We give the earliest war on our list index 1, the next index 2, and so on.

For the mean, we find the M_n statistic

$$M_{n,\text{mean}} = \max_{0.1 \leq s \leq 0.9} \left| \frac{\hat{\mu}_L - \hat{\mu}_R}{\{\hat{\sigma}_L^2/sn + \hat{\sigma}_R^2/n(1-s)\}^{1/2}} \right| = 1.93,$$

which has a p-value 0.50. For the standard deviation we find that

$$M_{n,\text{sd}} = \max_{0.1 \leq s \leq 0.9} \left| \frac{\hat{\sigma}_L - \hat{\sigma}_R}{\{\frac{\hat{\sigma}_L^2}{2ns} + \frac{\hat{\sigma}_R^2}{2n(1-s)}\}^{1/2}} \right| = 2.83,$$

which has a p-value 0.08.

On a 0.05 level we accept the null hypothesis of homogeneity, that the number of people killed in battle per war has followed the same trend from 1823 to 2003. On a 0.10 level, we reject that the standard deviation is homogeneous.

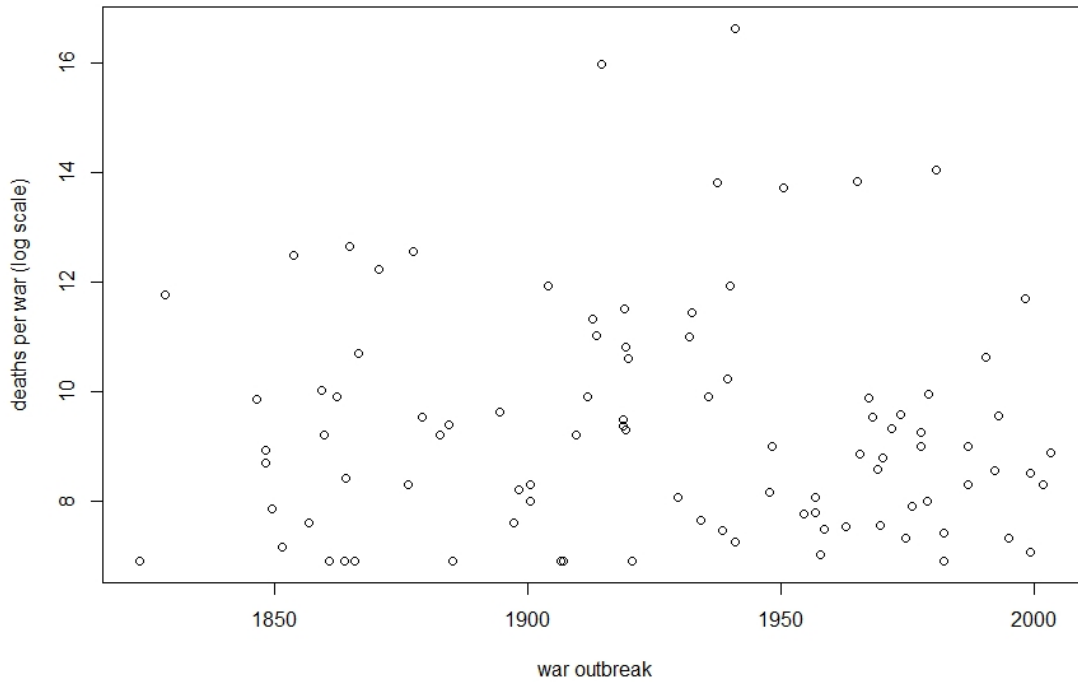


Figure 3.3: The outbreak of war and the logarithm $y_i = \log z_i$ of the number of battle deaths for 95 wars.

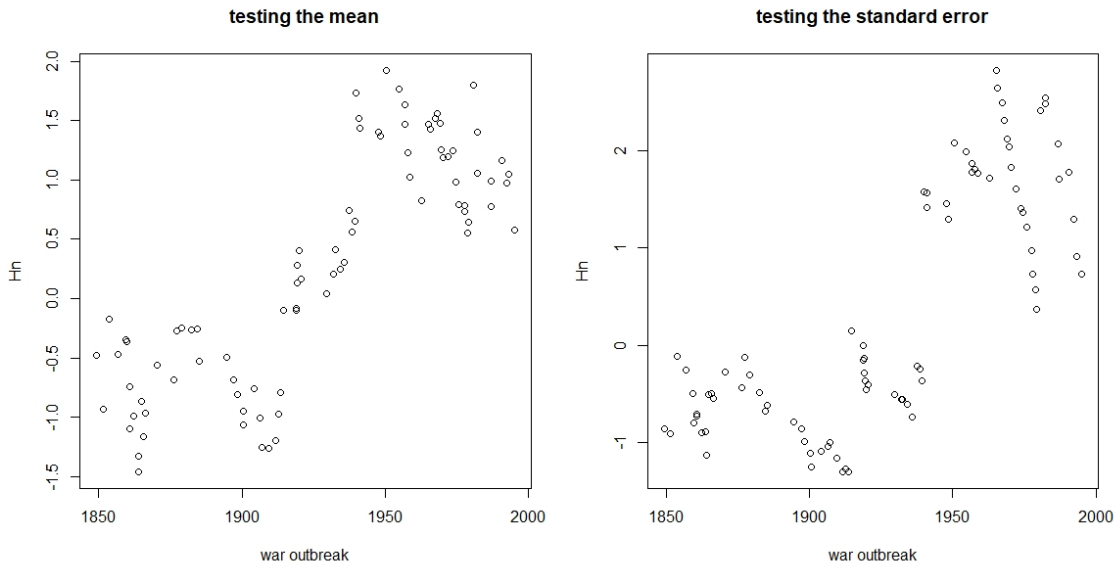


Figure 3.4: War outbreak (τ) and the monitoring process of the battle deaths data. Left: The $H_n(\tau, \bar{y})$ process of the mean. Right: The $H_n(\tau, \hat{\sigma})$ process of the standard deviation. The five first and last values of the process are not plotted, since we are interested in the process when there are at least a few observations on each side of τ .

Chapter 4

Power

In Chapter 3 we defined the test statistic

$$M_n = \max_{0.1 \leq s \leq 0.9} |H_n^*(s, T)| = \max_{0.1 \leq s \leq 0.9} \left| \frac{\hat{T}_L - \hat{T}_R}{\{\hat{\kappa}_L^2/ns + \hat{\kappa}_R^2/n(1-s)\}^{1/2}} \right|.$$

In this chapter we compare the hypothesis test using the test statistic M_n to other hypothesis tests. A hypothesis test will not always come to the correct conclusion, there are two types of errors a hypothesis test can make, denoted Type I and Type II.

Definition 4.1. A *Type I error* is when the test incorrectly rejects H_0 , when the null hypothesis is true. A *Type II error* is when the test incorrectly accepts H_0 , when the null hypothesis is false.

To estimate the Type II error of a test, we evaluate the power of the test. The power of a test is the probability of rejecting the null hypothesis when the alternative hypothesis is true. Ideally, we want the power function close to 1 when the alternative hypothesis is true, when $f(y, \theta) \in \mathcal{F}_A$, and close to 0 then the null hypothesis is true, when $f(y, \theta) \in \mathcal{F}_0$.

Definition 4.2. The power of a hypothesis test is its probability of rejecting the null hypothesis when the alternative hypothesis is true.

$$\Pr(\text{reject } H_0 | H_A \text{ is true}).$$

More formally, the power of a hypothesis test with rejection region R is the function of θ defined by $\beta(\theta) = \Pr_\theta(Y \in R)$.

We will compare the M-test to the likelihood ratio test and a test based on a Chi-squared statistic.

Theorem 4.3. [Casella and Berger, 2002] Let $Y_1, \dots, Y_n$ be a random sample from a pmf $f(x|\theta)$. The likelihood ratio statistic is defined as the maximum of the likelihood function of the null model divided by the maximum of the likelihood function of the alternative model

$$\lambda(y) = \frac{\sup_{\Theta_0} L(\theta|y)}{\sup_{\Theta} L(\theta|y)}$$

Under some regularity conditions, the distribution of the statistic $-2 \log \lambda(Y)$ converges to a Chi-squared distribution as the sample size increases. The degrees of freedom of the limiting distribution is the difference between the number of free parameters under the null hypothesis (with parameter space Θ_0), and the number of free parameters under the alternative hypothesis (with parameter space Θ).

When we use the likelihood ratio test to test the homogeneity of the mean of normally distributed observations, the test statistic $\lambda(y)$ has a Chi-squared distribution with one degree of

freedom. When all parameters are unknown, the test statistic is

$$\begin{aligned}\lambda(Y) &= \frac{\hat{\sigma}_0^{-n} \exp \left[-\sum_{i=1}^n (y_i - \hat{\mu}_0)^2 / 2\hat{\sigma}_0^2 \right]}{\hat{\sigma}_1^{-n} \exp \left[-\sum_{i=1}^{\tau} (y_i - \hat{\mu}_L)^2 / 2\hat{\sigma}_1^2 - \sum_{i=\tau+1}^n (y_i - \hat{\mu}_R)^2 / 2\hat{\sigma}_1^2 \right]} \\ &= \left(\frac{\hat{\sigma}_1}{\hat{\sigma}_0} \right)^n \exp \left(-\sum_{i=1}^n (y_i - \hat{\mu}_0)^2 / 2\hat{\sigma}_0^2 + \sum_{i=1}^{\tau} (y_i - \hat{\mu}_L)^2 / 2\hat{\sigma}_1^2 + \sum_{i=\tau+1}^n (y_i - \hat{\mu}_R)^2 / 2\hat{\sigma}_1^2 \right).\end{aligned}$$

We then find the Chi-squared statistic

$$-2 \log \lambda(Y) = -2n \ln(\hat{\sigma}_1 / \hat{\sigma}_0) + \sum_{i=1}^n (y_i - \hat{\mu}_0)^2 / \hat{\sigma}_0^2 - \sum_{i=1}^{\tau} (y_i - \hat{\mu}_L)^2 / \hat{\sigma}_1^2 - \sum_{i=\tau+1}^n (y_i - \hat{\mu}_R)^2 / \hat{\sigma}_1^2. \quad (4.1)$$

where $\hat{\mu}_0$ is the estimated mean and $\hat{\sigma}_0$ is the estimated standard deviation under the null hypothesis. The estimated standard deviation under the model with a change-point is $\hat{\sigma}_1$ and the estimated means are $\hat{\mu}_L$ and $\hat{\mu}_R$.

We now want to simulate power when the mean has a change-point. We let δ be a sequence of increasing values. For every value δ_j we let B be a large number and draw B data-sets from the normal distribution, where $y_1, \dots, y_{250} \sim N(0, 1)$ and $y_{251}, \dots, y_{500} \sim N(\delta_j, 1)$. We can then test the homogeneity of each data-set and observe how often the null hypothesis is rejected for each value of δ . Since the data we simulate has standard deviation equal to one, a different test statistic is the Chi-squared distributed statistic Q_n with $n - 1$ degrees of freedom, defined as

$$Q_n = \sum_{i=1}^n (y_i - \bar{y})^2 \sim \chi_{n-1}^2 \quad (4.2)$$

We simulate and compare the power of the likelihood ratio test, the test based on the statistic Q_n and the M-test. We let the likelihood ratio test have the unfair advantage of knowing the true change-point. hence we expect its power function to be above the the power function of the M-test. The Q_n statistic does not know the change-point, but knows the variance. All three test rejects the null hypothesis at level 0.05, which means that the probability of Type I error is at most 5%. We notice from Figure [4.1](#) that the M-test does much better than the test based on Q_n , compared to the likelihood ratio test.

We also, similarly as with the mean, simulate the power function for the likelihood ratio test and the M-test (Figure [4.2](#)) when there is a change-point in the standard deviation. We do this by drawing data-sets from the normal distribution where $y_1, \dots, y_{250} \sim N(0, 1)$ and $y_{251}, \dots, y_{500} \sim N(0, 1 + \delta_j)$.

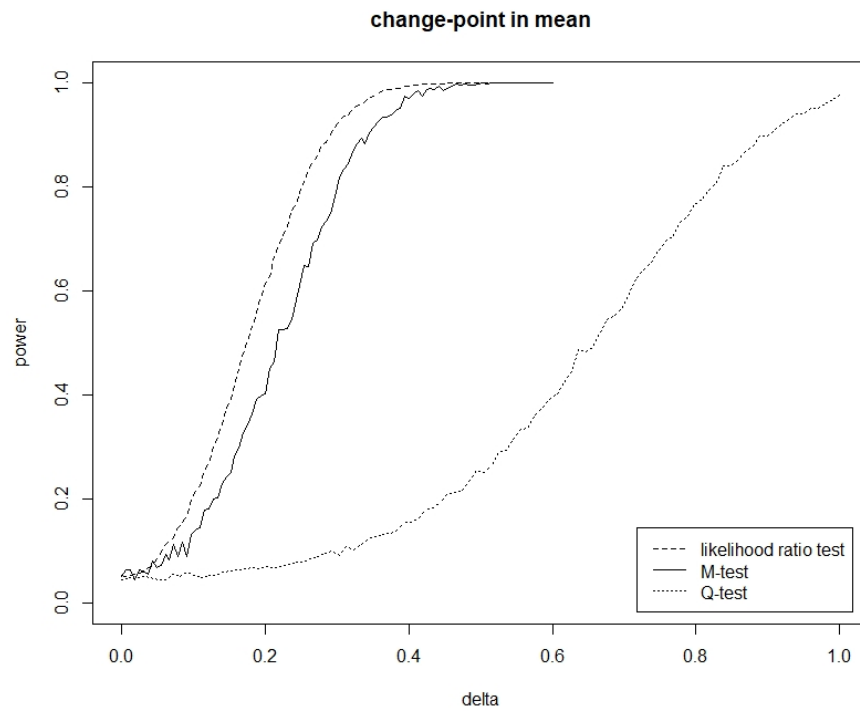


Figure 4.1: Power functions for three test as a function of the size of the shift δ at the change-point. We have the likelihood ratio test, the M-test and the test based on the statistic Q_n . All three tests reject the null hypothesis at level 0.05. The power functions are computed using simulated data-sets where $y_1, \dots, y_{250} \sim N(0, 1)$ and $y_{251}, \dots, y_{500} \sim N(\delta_j, 1)$.

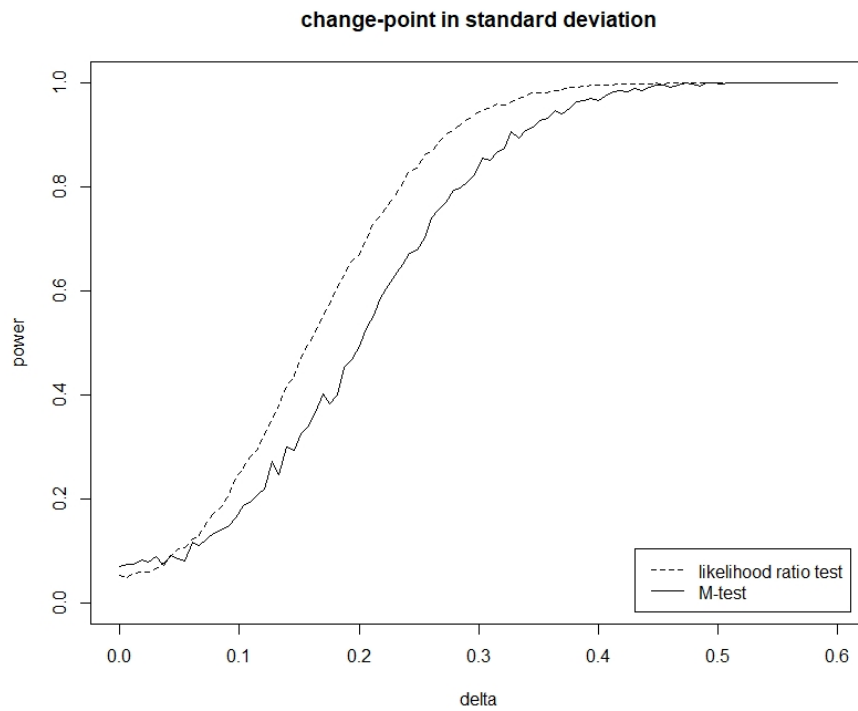


Figure 4.2: Power functions for two test as a function of the size of the shift δ at the change-point. We have the likelihood ratio test and the M-test. Both tests reject the null hypothesis at level 0.05. The power functions are computed using simulated data-sets where $y_1, \dots, y_{250} \sim N(0, 1)$ and $y_{251}, \dots, y_{500} \sim N(0, 1 + \delta_j)$.

Chapter 5

The behaviour of the monitoring process

In this chapter we look at how the choice of focus parameter T affects the monitoring process $H_n(\tau, T)$ depending on whether the hypothesis of homogeneity is true or false. We remind ourselves that the monitoring process is defined as

$$H_n(\tau, T) = \frac{\hat{T}_L - \hat{T}_R}{\{\hat{\kappa}_L^2/\tau + \hat{\kappa}_R^2/(n - \tau)\}^{1/2}}.$$

5.1 Parametric vs non-parametric tests

The assumptions we make about the observations affects how we define the monitoring process $H_n^*(s, T)$. The denominator of $H_n^*(s, T)$ depend on which assumptions we make about the distribution of the observations we are working with. It is an advantage of the M-test that we can construct the monitoring process without assuming a distribution.

5.2 When the hypothesis of homogeneity is true

We compare how the monitoring process $H_n(\tau, T)$ behaves depending on the focus parameter T when the null hypothesis of homogeneity is true. We simulate a standard normal data-set of 200 values, find and plot the monitoring process for the mean μ , the variance σ^2 , the standard deviation σ and the logarithm of the standard deviation $\log(\sigma)$. We plot the example in Figure 5.1. All the three processes involving the parameter σ are very similar, but we would have to simulate many cases in order to reach a conclusion about how similar we can expect the monitoring processes to be. In this case, for all the four processes, the test statistic M_n satisfy $M_n < 2.64$, where we accept the null hypothesis at a 0.1 level.

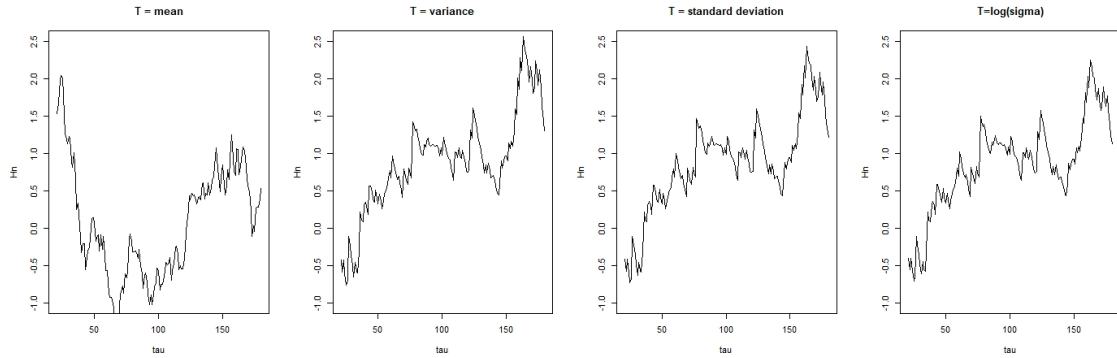


Figure 5.1: Example of the monitoring process under the null hypothesis. The focus parameter T is the mean μ , the variance σ^2 , the standard deviation σ , and the logarithm of the standard deviation $\log \sigma$. The processes are based on a sequence of 200 standard normal values.

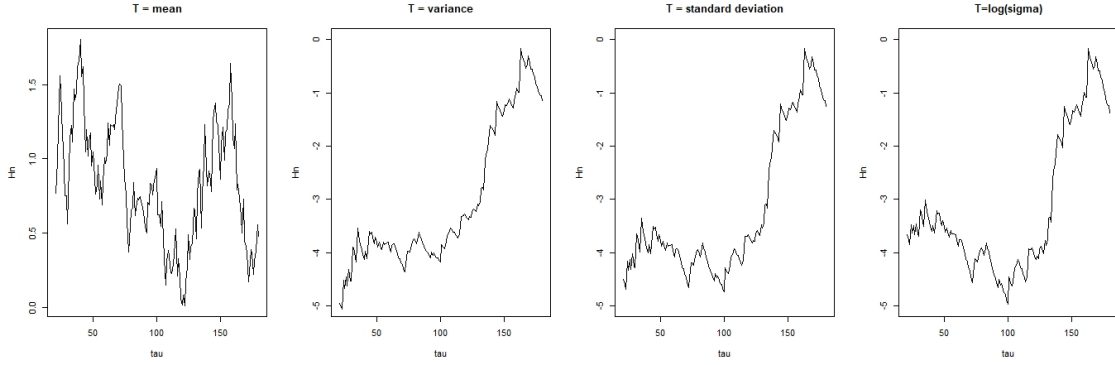


Figure 5.2: Example of the monitoring process when there is a change-point. The process is based on the sequence where $y_1, \dots, y_{100} \sim N(0, 1)$ and $y_{101}, \dots, y_{200} \sim N(0, 1.5)$. The focus parameter T is the mean μ , the variance σ^2 , the standard deviation σ , and the logarithm of the standard deviation $\log \sigma$.

5.3 When the hypothesis of homogeneity is false

We want to look at how the choice of focus parameter T affects the monitoring process when the null hypothesis is false and there is a change-point. We know that the process converges to a well defined Brownian motion process under the null hypothesis for most choices of focus parameter T , but we do not know the asymptotic behaviour of process when the null hypothesis is false.

We simulate a data-set with a change-point and look at how the process behaves. We simulate $y_1, \dots, y_{100} \sim N(0, 1)$ and $y_{101}, \dots, y_{200} \sim N(0, 1.5)$. We plot the processes for different focus parameters T , as in the previous section, in Figure 5.2. For all three focus parameters $T = \sigma$, $T = \sigma^2$ and $T = \log \sigma$, the M-test rejects H_0 at a 10^{-4} level. We see that the processes are less similar in terms of where the most extreme value occurs. For the variance, the smallest (and most extreme) value clearly is at the beginning of the process. For the standard deviation it is not as obvious when the process takes its smallest value, but the most extreme value is at the true change-point 100. For the logarithm of the standard deviation the process quite clearly has its most extreme value at the true change-point. This is an indication that when the null hypothesis is false, our choice of focus parameter T has a noteworthy influence on the monitoring process.

For our example, the p-value of the test statistic $M_n(\mu)$ of the mean is 0.41, which is not (and should not be) significant.

We now look at an example where the null hypothesis is much more obviously wrong, and see how the monitoring process behaves depending on T . We simulate a new data-set where $y_1, \dots, y_{100} \sim N(0, 1)$ and $y_{101}, \dots, y_{200} \sim N(0, 5)$, and plot the process where the focus parameter T is different functions of σ (Figure 5.3). Here, we see a similar pattern as in Figure 5.2. The process for the logarithm of the standard deviation $T = \log(\sigma)$ clearly has its most extreme value at the change-point. For the variance and the standard deviation, the processes start at its most extreme values.

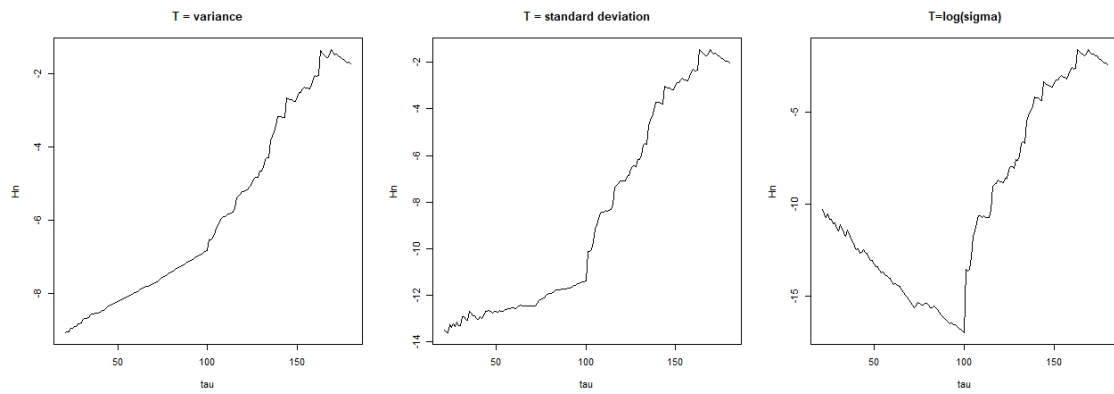


Figure 5.3: Example of the monitoring process when there is a change-point. The focus parameter T is the variance σ^2 , the standard deviation σ , and the logarithm of the standard deviation $\log \sigma$. The processes are based on a sequence where $y_1, \dots, y_{100} \sim N(0, 1)$ and $y_{101}, \dots, y_{200} \sim N(0, 5)$.

Chapter 6

Estimating the change-point

In this chapter we assume that we have rejected the the null hypothesis of homogeneity, and want to estimate a model with one change-points. We suggest an estimate for the change-point using our monitoring process.

We assume that we have observations from a model $f(y, \theta)$, and that there is one change-point τ such that

$$y_i \sim \begin{cases} f(y, \theta_L) & \text{if } i \leq \tau \\ f(y, \theta_R) & \text{if } i > \tau \end{cases}$$

where θ_L and θ_R are two different vectors of parameter values. In Chapter 3 we defined the test statistic

$$M_n = \max_{0.1 \leq s \leq 0.9} |H_n^*(s, T)| = \max_{0.1 \leq s \leq 0.9} \left| \frac{\hat{T}_L - \hat{T}_R}{\{\hat{\kappa}_L^2/ns + \hat{\kappa}_R^2/n(1-s)\}^{1/2}} \right|.$$

In the specific realization presented in Figure 5.2 when the null hypothesis was false, the index τ that corresponded to the statistic M_n was the true change-point for both $T = \log \sigma$ and $T = \sigma$. In Figure 5.3 the value of τ that corresponded to M_n was the true change-point for $T = \log \sigma$. A candidate for the estimated change-point is therefore the τ that correspond to M_n , lets call this the M-estimate. We see from the figures mentioned that this is not a good estimate in all cases, the plotted monitoring process for the variance and standard deviation in Figure 5.3 makes this clear. A maximum likelihood estimate is more robust in the sense that we can find an estimate $\hat{\tau}$ for the change-point for any model where the likelihood function exist. The strength of using M-test is that we do not necessarily have to assume a specific model.

When we estimate the change-point using maximum likelihood we use the profile log-likelihood function. Let $\ell_{1,\tau}(\hat{\theta}_L(\tau))$ be the likelihood function given the observations $y_1, \dots, y_\tau$, and let $\ell_{\tau+1,n}(\hat{\theta}_R(\tau))$ be the likelihood function given the observations $y_{\tau+1}, \dots, y_n$. The maximum likelihood estimate of the change-point is

$$\hat{\tau}_{ML} = \max_{\tau} \ell_{prof}(\tau) = \max_{\tau} \left\{ \ell_{1,\tau}(\hat{\theta}_L(\tau)) + \ell_{\tau+1,n}(\hat{\theta}_R(\tau)) \right\}$$

We investigate how accurate we can estimates a change-point using the M-estimate. We simulate a data-set where $y_1, \dots, y_{100} \sim N(0, 1)$ and $y_{101}, \dots, y_{200} \sim N(0.5, 1)$, and estimate the change-point of the mean. We repeat this 10^5 times to find the distribution of the estimate. We then simulate data-sets where $y_1, \dots, y_{100} \sim N(0, 1)$ and $y_{101}, \dots, y_{200} \sim N(0, 1.5)$, and similarly estimate the variance, standard deviation and the logarithm of the standard deviation, and find the distribution of these estimates. The four estimated distributions are plotted in Figure 6.1

We can now say something about the distribution of the M-estimate for the case where we have a normally distributed sequence of observations with one change-point. We notice that the distribution of the M-estimate is approximately symmetric around the true change-point when testing the mean and the logarithm of the standard deviation. For the standard deviation the M-estimate is slightly poorer, and is biased towards estimating the change-point as occurring earlier

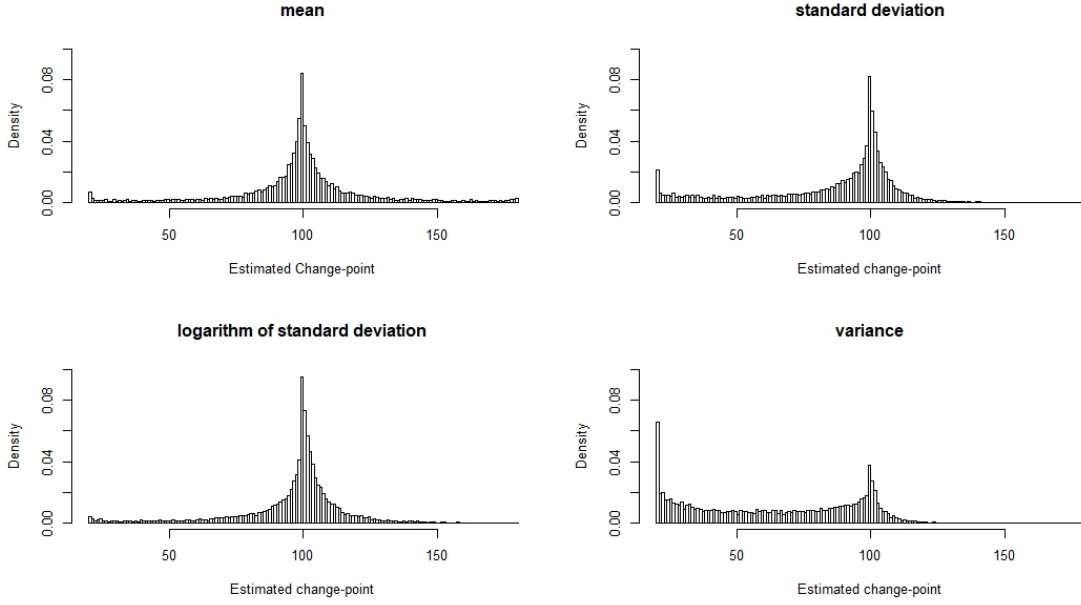


Figure 6.1: The distribution of the M-estimate of the change-point. We simulate sequences of 200 normally distributed values, where there is a change-point at index 100. The focus parameter increases with 0.5 at the change-point in all four cases. Top left: The estimated change-point of the mean. Top right: The estimated change-point of the standard deviation. Bottom left: The estimated change-point of the logarithm of the standard deviation. Bottom right: The estimated change-point of the variance.

than the true change-point. We see from the figure that the the M-estimate of the variance is not useful in this case. The estimate is all too often earlier than the true change-point.

We also simulate the distribution of the maximum likelihood estimate of the change-point. We first simulate 10^5 normally distributed data-sets with a change-point in the mean, as before we let $y_1, \dots, y_{100} \sim N(0, 1)$ and $y_{101}, \dots, y_{200} \sim N(0.5, 1)$. We find the estimated change-point for each data-set by maximizing the profile log-likelihood function for each data-set, letting the estimated mean differ before and after the change-point, but keeping the estimated standard deviation constant. Secondly we simulate 10^5 normally distributed data-sets with a change-point in the standard deviation, where $y_1, \dots, y_{100} \sim N(0, 1)$ and $y_{101}, \dots, y_{200} \sim N(0, 1.5)$. We maximize the profile log-likelihood function for each data-set letting the estimated standard deviation differ before and after the change-point, but keeping the estimated mean constant. The estimated distributions of estimated change-points are plotted in Figure 6.2

Comparing the M-estimate to the maximum likelihood estimate, the M-estimate does well when the focus parameter is the mean or the logarithm of the standard deviation (for normally distributed data).

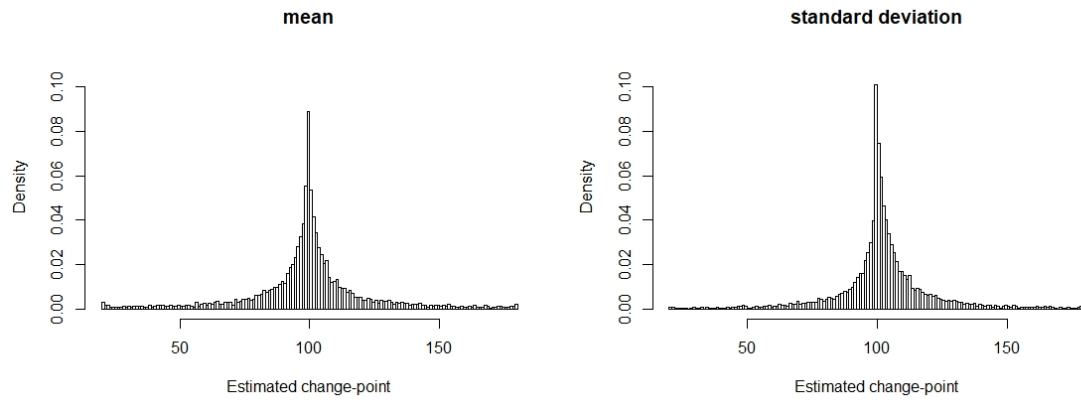


Figure 6.2: The distribution of the maximum likelihood estimate of the change-point. We simulated normally distributed data, where the true change-point is at index 100. The focus parameter increases with 0.5 at the change-point. Left: The estimated change-point of the mean. Right: The estimated change-point of the standard deviation.

Chapter 7

Confidence sets

In this chapter we assume that we have rejected the hypothesis of homogeneity, and assume that there is at least one change-point. We would like to assess the uncertainty around the estimated change-point $\hat{\tau}$, and want a confidence set which provides other probable values for the break-point. We use confidence curves to measure the uncertainty, there are several methods for constructing confidence distributions, two different methods are presented by [Cunen et al., 2018](#). From a confidence curve we can read off the confidence set for any given confidence level.

7.1 Confidence sets using profiled log-likelihood and deviance

We assume that there is one change-point and the observation stem from a parametric model $f(y, \theta)$. We estimate this change-point and denote the estimate $\hat{\tau}$. In this section we describe the confidence curve defined using profile log-likelihood and deviance.

Let $\hat{\theta}_L$ and $\hat{\theta}_R$ be the maximum likelihood estimates of the parameters on each side of an index τ , the observations $y_1, \dots, y_\tau$ are used to find $\hat{\theta}_L$ and $y_{\tau+1}, \dots, y_n$ are used to find $\hat{\theta}_R$. Let the random data-set Y be distributed $f(y, \theta_L)$ for $y_1, \dots, y_\tau$ and $f(y, \theta_R)$ for $y_{\tau+1}, \dots, y_n$. The parameters θ_L and θ_R are estimated using the observed sequence y_{obs} .

We first find the profile log-likelihood function, defined as

$$\ell_{\text{prof}}(\tau) = \ell_{1,\tau}(\hat{\theta}_L(\tau)) + \ell_{\tau+1,n}(\hat{\theta}_R(\tau)).$$

The deviance function is defined

$$D(\tau, Y) = 2(\ell_{\text{prof}}(\hat{\tau}) - \ell_{\text{prof}}(\tau)).$$

The confidence curve is defined as

$$cc(\tau, y_{\text{obs}}) = \Pr_{\tau}\{D(\tau, Y) < D(\tau, y_{\text{obs}}) | \hat{\theta}_L(\tau), \hat{\theta}_R(\tau)\}.$$

We estimate the confidence curve using a large number of simulated data-sets Y^* drawn from Y . Let B be a large number, the estimated confidence curve is

$$\hat{cc}(\tau, y_{\text{obs}}) = B^{-1} \sum_{i=1}^B I\{D(\tau, Y_i^*) < D(\tau, y_{\text{obs}})\} \quad (7.1)$$

For any desired confidence level α , the confidence set of $\hat{\tau}$ is

$$R(\alpha) = \{\tau : \hat{cc}(\tau, y_{\text{obs}}) \geq \alpha\} \quad (7.2)$$

By construction, the value of the confidence curve at the estimated change-point $\hat{\tau}$ is zero.

For illustration we simulate a sequence of Poisson distributed values with a change-point, letting $y_1, \dots, y_{50} \sim \text{Pois}(1)$ and $y_{51}, \dots, y_{100} \sim \text{Pois}(2)$. The maximum likelihood estimate $\hat{\tau}$ of the change-point is 47. We estimate and plot the confidence curve by letting $B = 300$ and plot the curve in Figure [7.1](#)

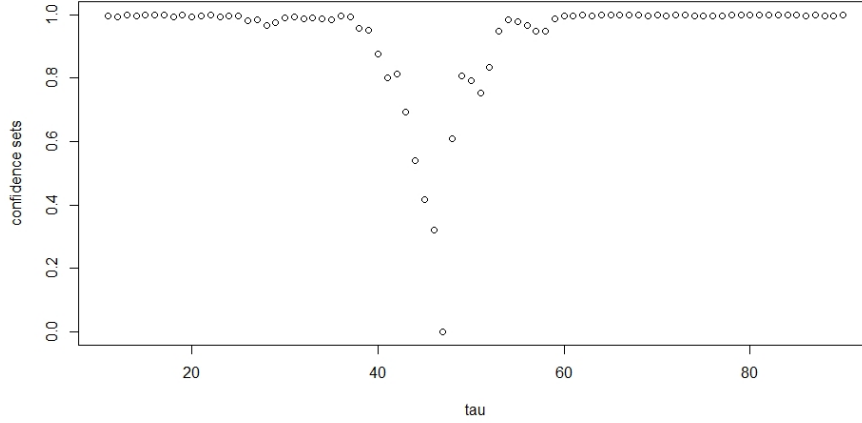


Figure 7.1: Estimated confidence curve using profile log-likelihood of the simulated sequence where $y_1, \dots, y_{50} \sim \text{Pois}(1)$ and $y_{51}, \dots, y_{100} \sim \text{Pois}(2)$, letting $B = 300$.

We check that the realized coverage is close to the specified confidence level for the confidence curve. We do this by simulating 500 data-sets Y_{obs} , where $y_1, \dots, y_{50} \sim \text{Pois}(1)$ and $y_{51}, \dots, y_{100} \sim \text{Pois}(2)$. For each data-set Y_{obs} we find the maximum likelihood estimate $\hat{\tau}$ of the change-point, and calculate its confidence curve using $B = 100$ (in Equation 7.1). We can then count how many times the true change-point is in the confidence set for different levels α , we present the result in Table 7.1

theoretical confidence	simulated coverage
0.99	0.990
0.95	0.960
0.90	0.898
0.80	0.818
0.70	0.734

Table 7.1: Simulated coverage of some confidence sets $R(\alpha)$ of the test using profile log-likelihood. To construct the confidence curves we simulated 500 data-sets where $y_1, \dots, y_{50} \sim \text{Pois}(1)$ and $y_{51}, \dots, y_{100} \sim \text{Pois}(2)$, and used $B = 100$ (Equation 7.1).

7.2 Constructing confidence sets using our test statistic M_n

We now construct confidence sets using our test statistic M_n . For every possible change-point τ in a data-set, we divide the data-set into two subsets, the first set Y_L , is observations with indexes from 1 to τ , the second set, Y_R , is observations with indexes from $\tau + 1$ to n . We then test whether the two subsets are homogeneous. We do not test for a change-point directly.

We use our statistic M_n

$$M_n = \max_{0.1 \leq s \leq 0.9} |H_n^*(s, T)|$$

to test homogeneity of each subset. Let $M_\tau(Y_L)$ be the statistic for Y_L , and $M_{n-\tau}(Y_R)$ be the statistic for Y_R . Let m_α be the $1 - \alpha$ upper quantile of the limit distribution of the statistic M_n for a given value α . Say we want confidence level α , since we test the homogeneity of Y_L and Y_R at the same time, we achieve our confidence level α by testing each subset at level $\sqrt{\alpha}$. We then have a confidence set

$$R(\alpha) = \{\tau : M_\tau(Y_L) \text{ and } M_{n-\tau}(Y_R) \text{ accepted at level } \sqrt{\alpha}\},$$

Theoretical confidence	Simulated coverage		
	200 observations	500 observations	1000 observations
0.99	0.975	0.984	0.987
0.95	0.911	0.947	0.944
0.90	0.852	0.892	0.880
0.80	0.749	0.796	0.795
0.70	0.649	0.704	0.695
0.20	0.231	0.212	0.217
0.10	0.128	0.107	0.117

Table 7.2: Approximated coverage of the confidence sets of the M-test. The true change-point is at $s = 0.5$. We simulated 1000 sets where $y_1, \dots, y_{ns} \sim N(0, 1)$, and $y_{ns+1}, \dots, y_n \sim N(1, 1)$ for $n = 200, 500$ and 1000 .

more formally

$$R(\alpha) = \{\tau : M_\tau(Y_L) \leq m_{\alpha^{1/2}}, M_{n-\tau}(Y_R) \leq m_{\alpha^{1/2}}\}. \quad (7.3)$$

We can then define the confidence curve

$$cc(\tau, y_{\text{obs}}) = \min\{\alpha : \tau \in R(\alpha)\}$$

If the null hypothesis is true and the whole data-set is homogeneous, any two subsets of the data-set will also be homogeneous. Then the confidence sets corresponding to low confidence level α will typically contain relatively many elements. On the other hand, if there are more than one change-point, we will never have homogeneity in both Y_L and Y_R at the same time, and the confidence sets with a low confidence level α will typically be empty. If there is one change-point, the confidence sets with a small confidence level α should contain just a few element and preferably only the true change-point, given that we have made the right assumptions about the underlying model.

For this method too, we investigate the realized coverage of the confidence sets. We let $s = 0.5$ and $n = 200, 500$ and 1000 , and simulate $A = 500$ data-sets, with $y_1, \dots, y_{ns} \sim N(0, 1)$ and $y_{ns+1}, \dots, y_n \sim N(1, 1)$. For every data-set we calculate the confidence curve, and observe which confidence sets contains the true change-point. We then estimate the realized coverage level of the confidence sets for a given α and a given number of observation n by taking the average of how often the true change-point is in the confidence curve,

$$\frac{1}{A} \sum_{j=1}^A I(\tau_{\text{true}} \in cc(\tau, Y_j)).$$

When we find M_τ and $M_{n-\tau}$, we make sure both subsequences Y_L and Y_R contain at least 10% of the observations. The simulated confidence values are presented in Table 7.2

We notice from Table 7.2 that when the number of observations is small the coverage level is smaller than the theoretical confidence. This is likely due to the fact that the confidence curve we investigate here rests on the assumption that the statistic M_n has converged to a Brownian motion process. When we calculate confidence curves, $M_\tau(Y_L)$ and $M_\tau(Y_R)$ will often be based on relatively few observations. Ideally we need a large amount of observations for both $M_\tau(Y_L)$ and $M_\tau(Y_R)$, so that it make sense to compare the statistics to the quantiles of M . We found indications in Chapter 3 Table 3.1, and Figure 3.1 that with a small number of observations n the statistic M_n become larger. Therefore, when we use this confidence curve construction, without a very large n , we get confidence values that are smaller than the theoretical confidence of M . This is particularly a weakness of the test when there is a change-point close to the beginning or the end of a sequence of observations, in those cases we might not find the change-point. since $M_\tau(Y_L)$ or $M_\tau(Y_R)$ will be based on few observations and therefore have a larger expected value. When either $M_\tau(Y_L)$ or $M_\tau(Y_R)$ is large the null hypothesis of homogeneity is typically rejected and the confidence sets for τ will often be empty.

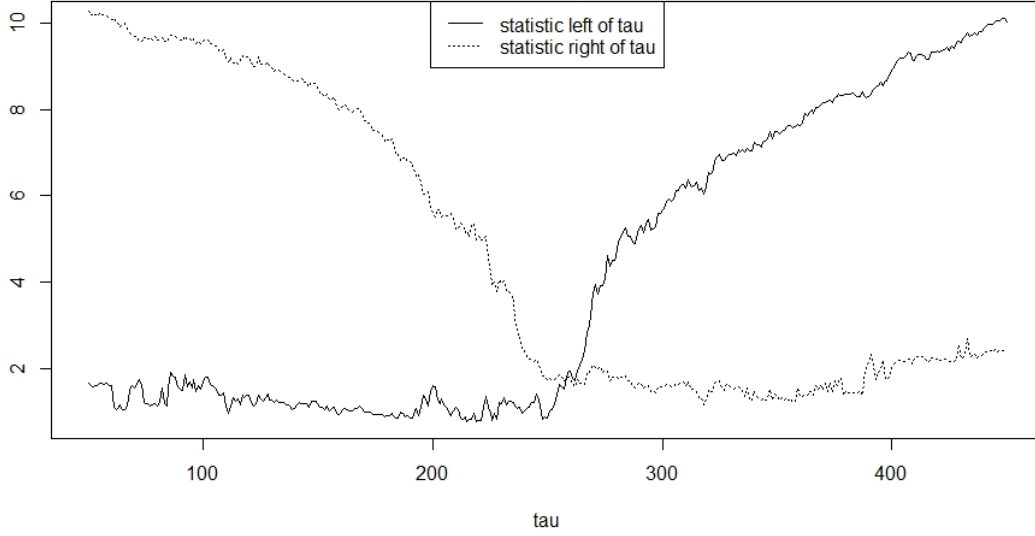


Figure 7.2: The solid line is $M_\tau(Y_L)$ and the dotted line is $M_{n-\tau}(Y_R)$ of a sequence where $y_1, \dots, y_{250} \sim N(0, 1)$ and $y_{251}, \dots, y_{500} \sim N(1, 1)$

To illustrate the test statistics $M_\tau(Y_L)$ and $M_{n-\tau}(Y_R)$ we simulated a data-set where $y_1, \dots, y_{250} \sim N(0, 1)$ and $y_{251}, \dots, y_{500} \sim N(1, 1)$ and plot $M_\tau(Y_L)$ and $M_{n-\tau}(Y_R)$ for τ between 50 and 450 (Figure 7.2). We see from the figure that both $M_\tau(Y_L)$ and $M_{n-\tau}(Y_R)$ are relatively small when $\tau = 250$, when it is true that Y_L and Y_R are homogeneous.

We illustrate the difference between using profile log-likelihood and using the M-test to make a confidence curve. We find and plot the confidence curve using the statistic M_n for the same data-set we used in Figure 7.1 in the previous section (the left plot of Figure 7.3).

Sine we suspect that the confidence set based on the M-test contains too many values for a given α when we only have 100 observations, we also simulate a sequence of 500 Poisson distributed values where $y_1, \dots, y_{250} \sim \text{Pois}(1)$ and $y_{251}, \dots, y_{500} \sim \text{Pois}(2)$, and find its confidence curve (the right plot in Figure 7.3).

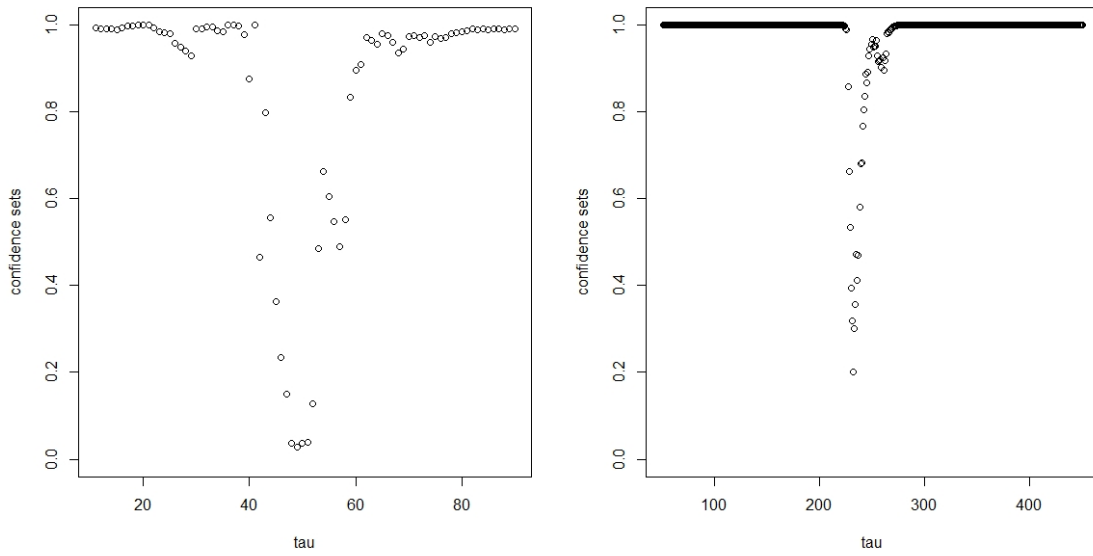


Figure 7.3: Confidence curves based on the M-test of sequences where $y_1, \dots, y_{0.5n} \sim \text{Pois}(1)$ and $y_{0.5n+1}, \dots, y_n \sim \text{Pois}(2)$. Left: The confidence curve of the same sequence of $n = 100$ values as used to make the confidence curve in Figure [7.1](#). Right: The confidence curve of a sequence with $n = 500$ values.

Chapter 8

Application: British coal-mining disasters

We look at the number of British coal-mining disasters per year, for 111 years, from 1851 to 1962. The disasters are 191 explosions involving 10 or more men killed. For more information about the data-set see [Jarrett, 1979](#). This is a much analyzed data-set, and is used as an example by [Cunen et al., 2018](#). We will investigate whether the disasters are distributed equally over time and estimate the potential change-points.

First we will assume that the observations are normally distributed and test the homogeneity of the mean and the variance. Secondly we assume that the observations has a Poisson distribution and test the homogeneity of the Poisson rate. The null hypothesis of homogeneity is rejected, and we find indication of the same two change-points in both cases. We thirdly assume that the observations stem from a Gamma-Poisson distribution, and find the most likely change-points for this case as well.

In the article by [Cunen et al., 2018](#), the observations are assumed to be independent and Poisson distributed. The article rejected the hypothesis of homogeneity, and assumes that there is a single change-point. They find that the maximum likelihood estimate of the change-point is $\hat{\tau} = 1891$, and find a confidence set for the observations by profile log-likelihood. In addition to 1891, we find that 1947 is a likely change-point.

Year	Count	Year	Count	Year	Count	Year	Count	Year	Count	Year	Count
1851	4	1871	5	1891	2	1911	0	1931	3	1951	1
1852	5	1872	3	1892	1	1912	1	1932	3	1952	0
1853	4	1873	1	1893	1	1913	1	1933	1	1953	0
1854	1	1874	4	1894	1	1914	1	1934	1	1954	0
1855	0	1875	4	1895	1	1915	0	1935	2	1955	0
1856	4	1876	1	1896	3	1916	1	1936	1	1956	0
1857	3	1877	5	1897	0	1917	0	1937	1	1957	1
1858	4	1878	5	1898	0	1918	1	1938	1	1958	0
1859	0	1879	3	1899	1	1919	0	1939	1	1959	0
1860	6	1880	4	1900	0	1920	0	1940	2	1960	1
1861	3	1881	2	1901	1	1921	0	1941	4	1961	0
1862	3	1882	5	1902	1	1922	2	1942	2	1962	1
1863	4	1883	2	1903	0	1923	1	1943	0		
1864	0	1884	2	1904	0	1924	0	1944	0		
1865	2	1885	3	1905	3	1925	0	1945	0		
1866	6	1886	4	1906	1	1926	0	1946	1		
1867	3	1887	2	1907	0	1927	1	1947	4		
1868	3	1888	1	1908	3	1928	1	1948	0		
1869	5	1889	3	1909	2	1929	0	1949	0		
1870	4	1890	2	1910	2	1930	2	1950	0		

Figure 8.1: The number of coal-mining disasters per year, table taken from [Moreno et al., 2005](#)

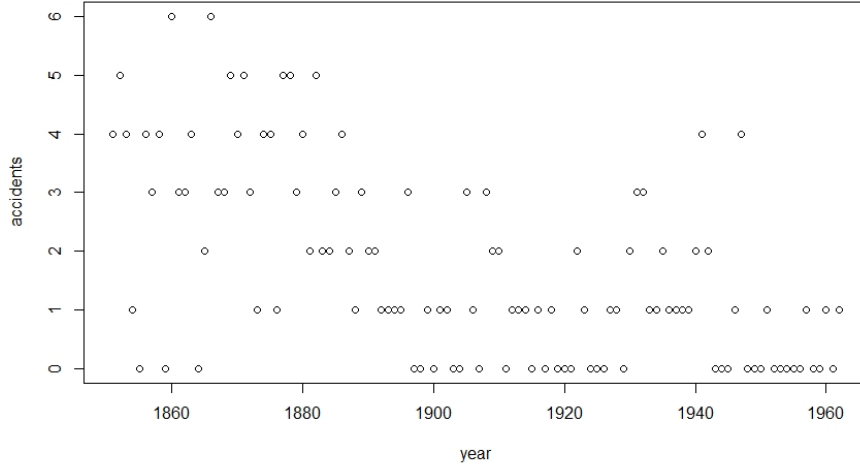


Figure 8.2: British coal-mining disasters per year

8.1 An analysis assuming normally distributed observations

In this section we assume that the observations are normally distributed, and test the homogeneity of the mean and the standard deviation.

Let $\hat{\sigma}$ be the estimated standard deviation and $\hat{\mu}$ be the estimated mean. We test the homogeneity of the mean using the monitoring process

$$H_n(s, \hat{\mu}) = \frac{\hat{\mu}_L - \hat{\mu}_R}{\{\hat{\sigma}_L^2/\tau + \hat{\sigma}_R^2/(n - \tau)\}^{1/2}},$$

and test the homogeneity of the standard deviation using the processes

$$H_n(s, \hat{\sigma}) = \frac{\hat{\sigma}_L - \hat{\sigma}_R}{\{\hat{\sigma}_L^2/2ns + \hat{\sigma}_R^2/2n(1 - s)\}^{1/2}}.$$

The monitoring processes are plotted in Figure 8.3. For the mean number of accidents per year, we find the statistic

$$M_{n,\text{mean}} = \max_{0.1 < s < 0.9} |H_n(\tau, \hat{\mu})| = 8.09$$

and for the standard deviation, we find the statistic

$$M_{n,\text{sd}} = \max_{0.1 < s < 0.9} |H_n(\tau, \hat{\sigma})| = 8.24.$$

When we simulated the limiting distribution of the statistic, the largest value we drew was 5.32, therefore, the p-value of $M_{n,\text{mean}}$ and $M_{n,\text{sd}}$ is less than 10^{-5} , so we reject the hypothesis of homogeneity for both the mean and the standard deviation.

Both the maximum of $H_n(s, \hat{\mu})$ and the maximum of $H_n(s, \hat{\sigma})$ are found in year 1947. This indicates a change-point in 1947. From the table in Figure 8.1 we see that there was at most one accident per year after 1947. The shape of the H_n processes in Figure 8.3 especially for the mean, indicates that there are two change-points, since the process reaches two clear local maximums. There might be a break-point both in 1891 (the break-point found by Cunen et al., 2018) and in 1947.

Lets assume that there is a change-point in 1891. We test the homogeneity of the observations after 1891, to further investigate whether a second change-point is likely. We plot the same H_n processes as in Figure 8.3 again in Figure 8.4 this time only considering observations after 1891.

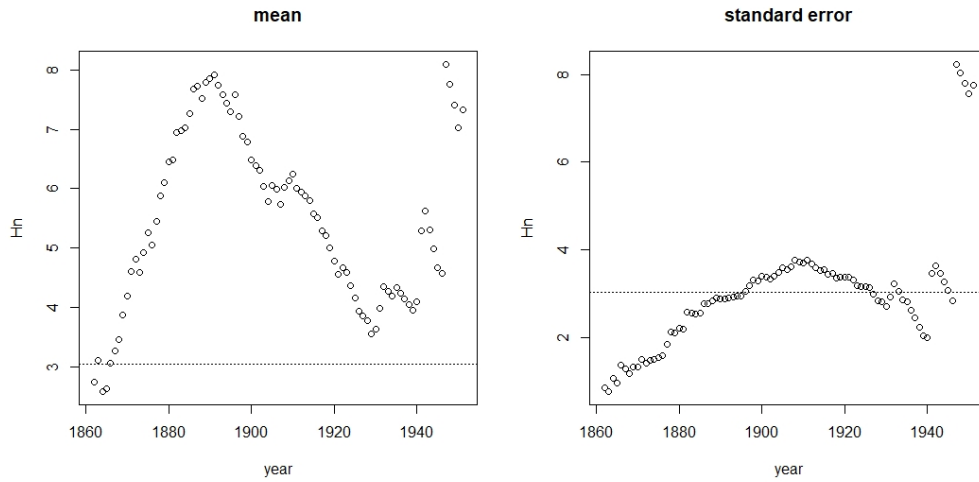


Figure 8.3: The monitoring process assuming the observations are normally distributed. Left: The process of the mean, Right: The process of the standard deviation. The first and last 11 years are not plotted but are included in the calculations, corresponding to letting $0.1 < s < 0.9$. The dotted line is the 0.95 quantile of the limit distribution of the statistic M_n . We reject the null hypothesis since the process takes large values.



Figure 8.4: The monitoring process when we only consider the observations after year 1891. We assume that the observations are normally distributed. Left: The process of the mean, Right: The process of the standard deviation. The first and last 7 years of the sequence, starting at 1892, are not plotted but are included in the calculations, corresponding to letting $0.1 < s < 0.9$. The dotted line is the 0.95 quantile of the limit distribution of the statistic M_n . We reject the hypothesis of homogeneity since the process takes values above the line.

We notice that the statistics $M_{n,\text{mean}}$ and $M_{n,\text{sd}}$ are still significant, but not as significant as before. For both the mean and the standard deviation we again reject the null hypothesis at a 0.05 level and the processes still has their maximum values in year 1947. This indicates a change-point in 1947, in addition to 1891.

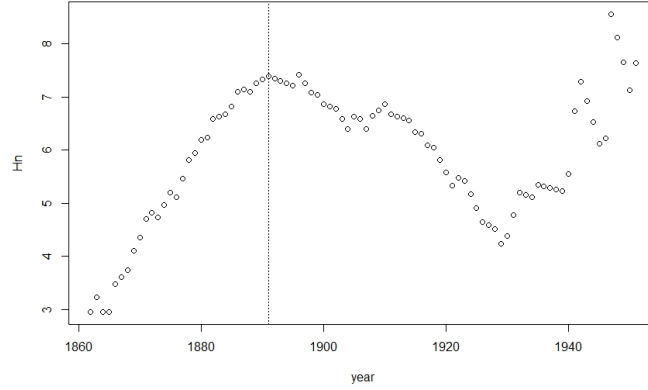


Figure 8.5: The monitoring process for the disasters, assuming a Poisson distribution with rate λ . The dotted line is 1891, the maximum likelihood estimate of the change-point. The process takes large values and rejects the hypothesis of homogeneity for all reasonable confidence levels.

8.2 Assuming a Poisson distribution

If we assume that accidents occur independently of the time since the last accident, it is reasonable to assume that the number of accidents per year follow a Poisson distribution with mean rate λ_i . Say we trust that the observations come from a Poisson distribution with density

$$f(y, \lambda) = \frac{\lambda^y e^{-\lambda}}{y!}.$$

Then, our monitoring process becomes

$$H_n^*(s, \lambda) = \frac{\hat{\lambda}_L - \hat{\lambda}_R}{\{\hat{\lambda}_L/sn + \hat{\lambda}_R/n(1-s)\}^{1/2}}.$$

The test statistic of the Poisson rate is $M_{n,\text{rate}} = 8.56$, which means that the p-value is less than 10^{-5} . Hence, we reject the hypothesis of homogeneity for all reasonable confidence levels. The shape of the H_n process (Figure 8.5) supports the hypothesis that there are two change-points since it has two clear local maximums.

We assume that there is at least one change-point. We now want to estimate the change-point(s). We let $0.1 \leq s \leq 0.9$. The index corresponding to $M_{n,\text{rate}} = 8.56$ is in 1947 (the 97th observation), which is the same year as the change-point we suggested in the previous section. If we ignore the first and last 20 observations, the most extreme value of H_n is 1896 (observation 46), which is relatively close to the 1891, the maximum likelihood estimate of the change-point found by [Cunen et al., 2018](#). The maximum likelihood estimate is marked with a dotted line in Figure 8.5. For the maximum likelihood estimate of the change-point, the estimated rates before and after 1891 are $\hat{\lambda}_L = 3.1$ and $\hat{\lambda}_R = 0.9$.

The right plot in Figure [8.6](#) shows the confidence curve bases on the M-test. For the M-test, almost all confidence sets except the set with confidence $\alpha = 1$ are empty. This indicates that we never have homogeneity on both sides of any one potential change-point under the Poisson assumption. We also plot $M_\tau(Y_L)$ and $M_{n-\tau}(Y_R)$ (defined the same way as in Equation [7.3](#)) in Figure [8.7](#), the statistic M_n for the data on each side of all potential change-points. We see that the maximum of $M_\tau(Y_L)$ and $M_{n-\tau}(Y_R)$ are larger than the 0.99 quantile of the distribution of M for all values of τ , which is 3.57. We therefore do not believe that we can divide the sequence of observations into two homogeneous subsequences. Nevertheless, We notice that $M_\tau(Y_L)$ is small, and less than the 0.5 quantile, from 1866 to 1891, which indicates that the Poisson rate is constant in this time interval. We have indication that there is another change-point after 1891.

The left plot in Figure [8.6](#) is the confidence curve based on profile log-likelihood, for the maximum likelihood estimate of the change-point 1891. This is under the assumption that there is

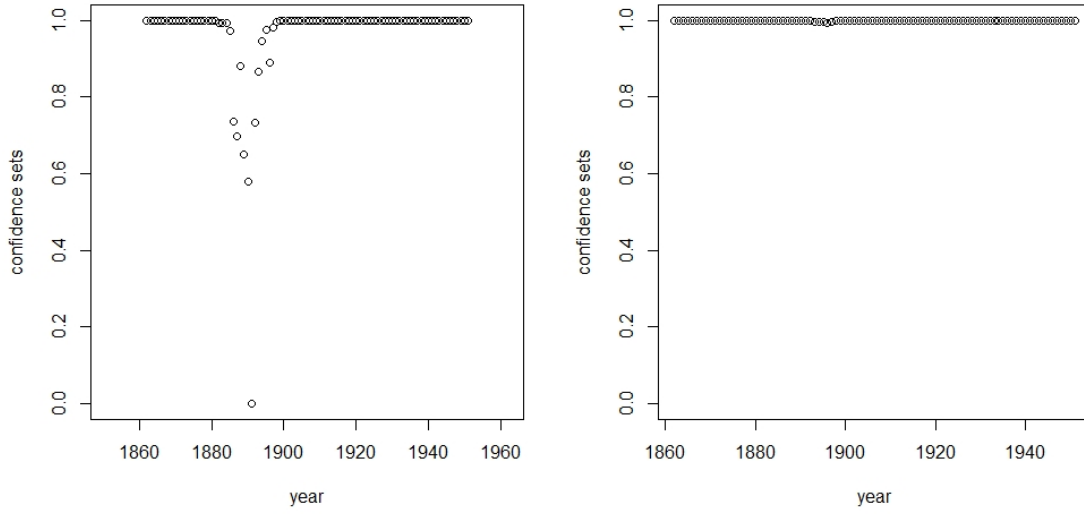


Figure 8.6: Assuming a Poisson distribution. Left: Confidence curves found by maximum likelihood for the estimated change-point 1891. Right: Confidence curves found by the M_n statistic, virtually no confidence is given to any one particular change-point.

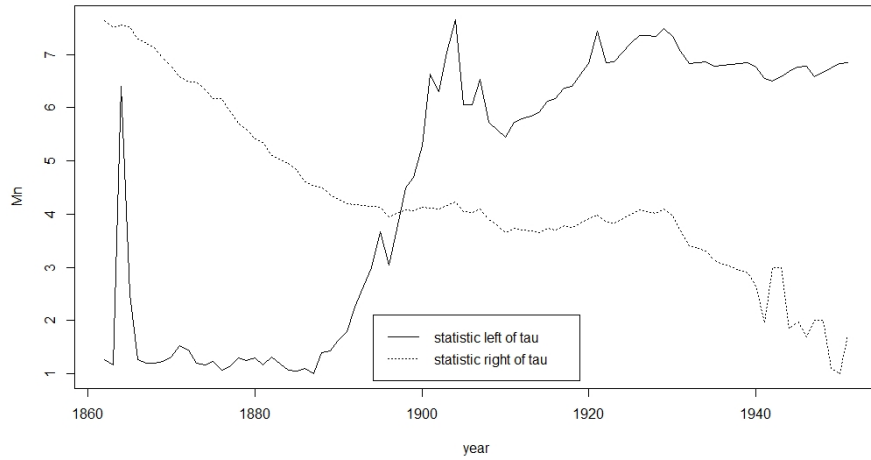


Figure 8.7: For year τ , the solid line is the the statistic M_τ of the observations from 1851 to τ , the dotted line is the statistic $M_{n-\tau}$ of the observations from year $\tau + 1$ to 1962. We never have two small values at the same time and hence do not believe that both the sequence of observations $y_1, \dots, y_\tau$ and the sequence $y_{\tau+1}, \dots, y_n$ are homogeneous for any given τ .

exactly one change-point. We see from the curve that most confidence sets with a given confidence level α contain relatively few elements, which indicated certainty.

We assume that we have two change-point, $\hat{\tau}_1 = 1891$ and $\hat{\tau}_2 = 1947$. We use the M-test to find confidence curves for the observations before 1947 and the observations after 1891 separately. We plot the confidence curves in Figure [8.8](#). The confidence curves indicate homogeneity between the change-points.

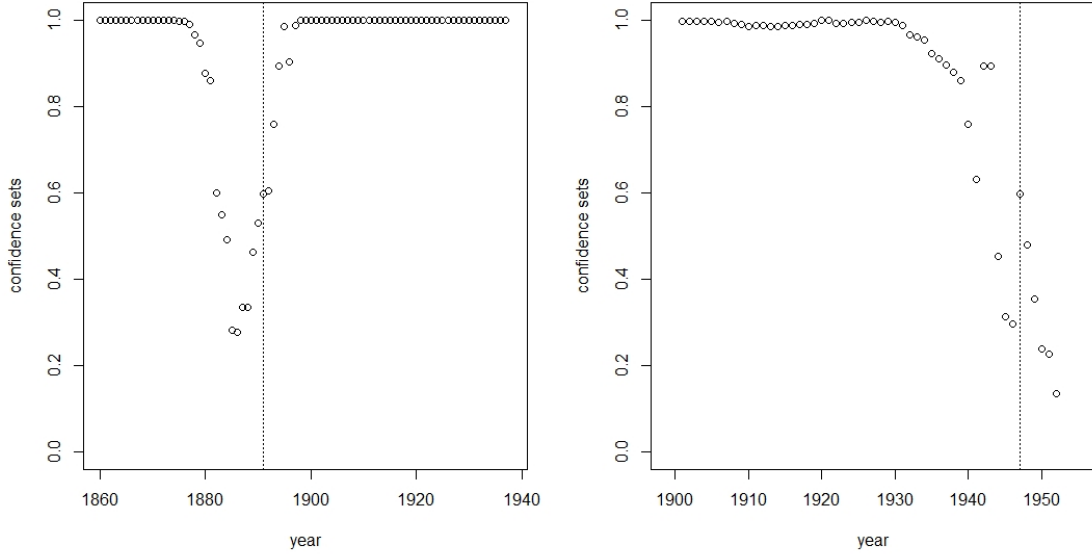


Figure 8.8: Assuming a Poisson distribution. Confidence curves found by the M_n statistic. Left: Confidence curve for the observations before 1947 Right: Confidence curve for the observations after 1891.

8.3 Assuming observations from a Gamma-Poisson distribution

We now extend the Poisson model to a Gamma-Poisson model, and assume that the observations stem from a Gamma-Poisson distribution, with parameters a_L and b_L for $y_1, \dots, y_\tau$, and parameters a_R and b_R for $y_{\tau+1}, \dots, y_n$.

Let the parameter θ_i have a Gamma distribution with density g , with scale parameter a_i and shape parameter b_i . When θ_i is given, the observations y_i have a Poisson distribution with rate θ_i . We denote the density of the Poisson distribution h ,

$$\begin{aligned} y|\theta &\sim \text{Pois}(\theta) \\ \theta &\sim \text{Gamma}(a, b). \end{aligned}$$

The density of the Gamma distribution is

$$g(\theta, a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}.$$

We find the density of y by integrating the joint distribution of y given θ and the random θ over all possible values of θ ,

$$\begin{aligned} f(y, a, b) &= \int_0^\infty h(y|\theta) g(\theta, a, b) d\theta = \int_0^\infty \frac{\theta^y e^{-\theta}}{y!} \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} d\theta, \quad y = 0, 1, 2, \dots \\ &= \frac{b^a}{\Gamma(a)y!} \int_0^\infty \theta^{y+a-1} e^{-(1+b)\theta} d\theta \\ &= \frac{b^a}{(b+1)^{a+y} \Gamma(a)y!} \Gamma(a+y). \end{aligned}$$

We denote the expected value of the observation $\lambda_i = E(Y_i)$, we then have

$$\lambda_i = a_i/b_i$$

model	Parameters					
1	$\hat{a} = 3.89$	$\hat{b}_L = 2.02$	$\hat{b}_R = 14.58$		$\lambda_L = 1.93$	$\lambda_R = 0.27$
2	$\hat{a}_L = 4.39$	$\hat{a}_R = 0.74$	$\hat{b} = 2.29$		$\lambda_L = 1.92$	$\lambda_R = 0.32$
3	$\hat{a}_L = 3.77$	$\hat{a}_R = 14687$	$\hat{b}_L = 1.95$	$\hat{b}_R = 55050$	$\lambda_L = 1.93$	$\lambda_R = 0.27$

Table 8.1: Estimated parameter values for Gamma-Poisson models with an estimated change-point in year 1947. Model 1: a is homogeneous and b is allowed to change in 1947, Model 2: b is homogeneous and a is allowed to change in 1947, Model 3: Both a and b are allowed to change in 1947.

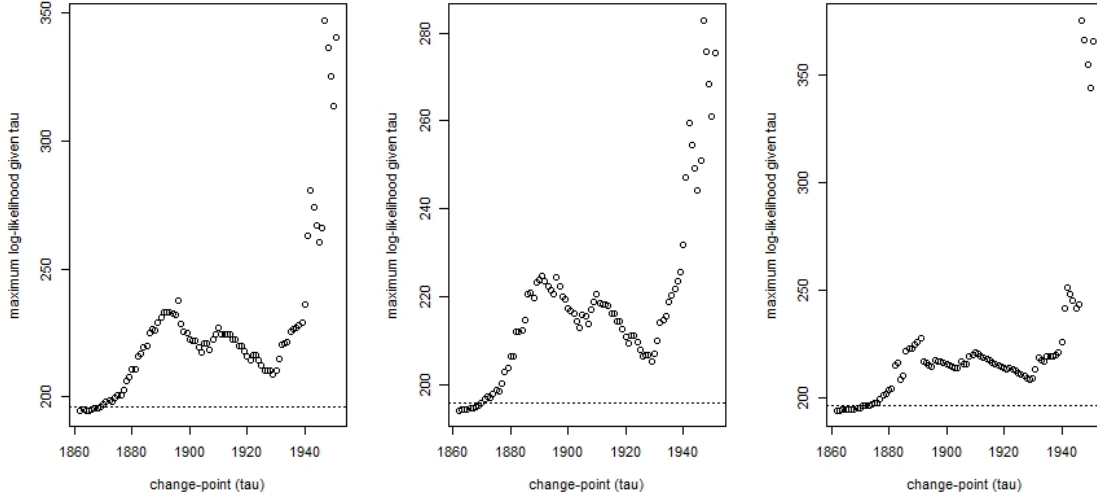


Figure 8.9: Maximum log-likelihood of the Gamma-Poisson model. Left: change-point in b , Middle: change-point in a , Right: both a and b are allowed to change at the change-point. The dotted line marks the maximum log-likelihood under the null hypothesis of homogeneity.

We can maximize the profile log-likelihood function using R, with the function `dgpoids` from the library `extraDistr` and the function `nlm`. We look at three different change-point models. For the first model, we let the parameter a of the Gamma-Poisson distribution be homogeneous and allowed b to change at the change-point, such that $y_i \sim f(y, a, b_L)$ for $i = 1, \dots, \tau$, and $y_i \sim f(y, a, b_R)$ for $i = \tau + 1, \dots, n$. We next allowed a to change at the change-point, and let b be homogeneous. Lastly we allow both a and b to change at the change-point. We do not consider a change-point in the first or last 11 observations (the first and last 10% of the sequence). We plot the log-likelihood function for all three cases in Figure 8.9. In all three cases, the maximum likelihood estimate of the change-point is year 1947. The maximum likelihood estimates of the parameters of the three model are presented in Table 8.1. Which gives us further evidence of a change-point in 1947.

Chapter 9

Application: Tirant lo Blanch

We will analyse the novel *Tirant lo Blanch* published in Valencia in 1490. Martorell wrote most of the novel, but died before finishing it. Many scholars agree that Martí Joan de Galba finished the novel and wrote the last part of it. For more see [Girón et al., 2005](#). We will do an analysis to find where Galba might have started writing. Since the novel has 487 chapters, we compare the chapters, and look for the chapter where Galba started writing. One author might have used long words more frequently than the other, hence we look at the proportion of words with a given length per chapter.

The number of words per chapter with a specific number of letters, say for example words with only one letter, has a binomial distribution. We look at the 425 chapters with more than 200 words. We look at the proportion of short words, which we define as words with 1, 2 or 3 letters, the proportion of words with medium length, which we define as words with 4,5 or 6 letters, and long words which we define as words with more that 7 letters. The proportions are plotted in Figure [9.1](#)

Let m_i be the total number of words in chapter i . We use j to indicate word length, and give short words index 1, medium length words index 2 and long words index 3. Let $x_{i,j}$ be the number of words in category j in chapter i . For an unknown probability $p_{i,j}$, the count $x_{i,j}$ has a binomial distribution

$$x_{i,j} \sim \text{Bin}(m_i, p_{i,j}).$$

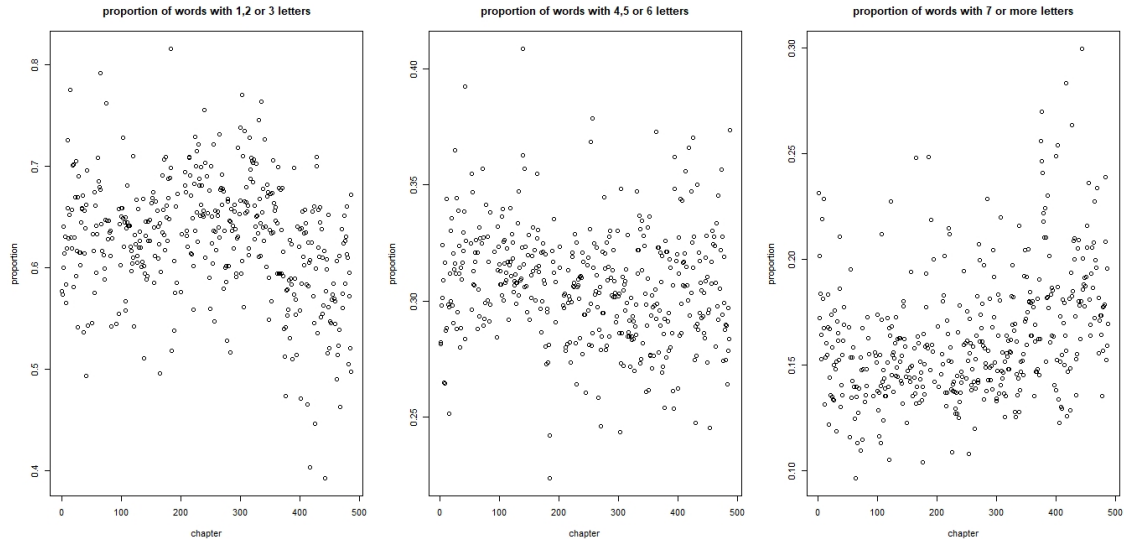


Figure 9.1: Left: The proportion of short words (words with 1,2 or 3 letters) per chapter. Middle: The proportion of words with medium length (words with 4,5 or 6 letters) per chapter. Right: The proportion of long words (words with 7 or more letters) per chapter.

We assume that the counts $x_{i,j}$ are independent. Since the chapters have different length, counts $x_{i,j}$ are not identically distributed and the shorter chapters have more variance. We are outside the iid assumption of chapter 2, and need more than Theorem 2.9 to state that the process H_n will converge to a Brownian motion process, but we have good reason to think that the theory will hold in this case, and will treat this case thereafter.

The estimate of the probability $p_{i,j}$ is

$$\hat{p}_{i,j} = x_{i,j}/m_i,$$

the proportion of words with length j per chapter. Our null hypothesis is that the probability of words with length j is the same in all chapters,

$$H_0 : p_{1,j} = p_{2,j} = \dots = p_{487,j}.$$

Let m_L be the total number of words from chapter 1 through chapter τ , and let m_R be the total number of words after chapter τ through the last chapter,

$$m_L = \sum_{i=1}^{\tau} m_i, \quad \text{and} \quad m_R = \sum_{i=\tau}^n m_i.$$

Define p_{Lj} as the proportion of words with length j from the beginning up to and including chapter τ . Define p_{Rj} as the proportion of words with length j after chapter τ ,

$$\hat{p}_{Lj} = \frac{\sum_{i=1}^{\tau} x_{i,j}}{m_L}, \quad \text{and} \quad \hat{p}_{Rj} = \frac{\sum_{i=\tau+1}^n x_{i,j}}{m_R}.$$

The variance of $\hat{p}_{Lj}$ is

$$\hat{\sigma}_{Lj}^2 = \frac{\sum_{i=1}^{\tau} \text{Var}(x_{i,j})}{m_L^2} = \frac{\sum_{i=1}^{\tau} m_i p_{i,j} (1 - p_{i,j})}{m_L^2}.$$

Similarly, the variance of $\hat{p}_{Rj}$ is

$$\hat{\sigma}_{Rj}^2 = \frac{\sum_{i=\tau+1}^n m_i p_{i,j} (1 - p_{i,j})}{m_R^2}.$$

We can then define the monitoring process

$$H_n(\tau, p_j) = \frac{\hat{p}_{Lj} - \hat{p}_{Rj}}{(\hat{\sigma}_{Lj}^2 + \hat{\sigma}_{Rj}^2)^{1/2}}.$$

The monitoring processes for the three categories word length are plotted in Figure 9.2. We reject the null hypothesis of homogeneity for all reasonable confidence levels.

For the short words, the monitoring process takes its most extreme value at chapter 371. For the medium length words, the monitoring process takes its most extreme value at chapter 167. For the long words, the monitoring process takes its most extreme value at chapter 345. Most researchers favouring the hypothesis of two authors agree that the change happened between chapter 350 and chapter 400. Our best guess, without further analysis, is therefore chapter 371. This is the same chapter as the maximum likelihood estimate found by Cunen et al., 2018.

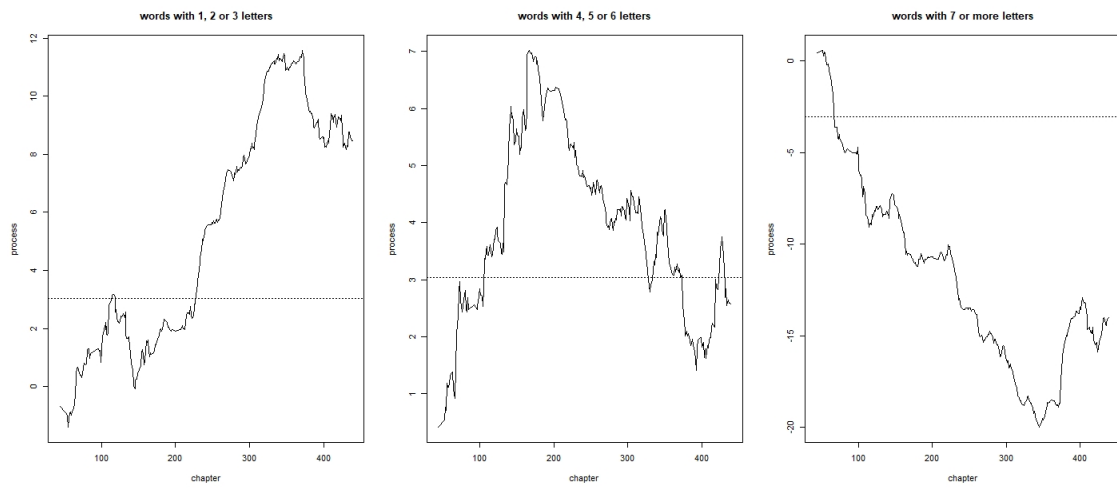


Figure 9.2: Left: The H_n process for the proportion of short words (words with 1,2 or 3 letters) per chapter. Middle: The H_n process for the proportion of words with medium length (words with 4,5 or 6 letters) per chapter. Right: The H_n process for the proportion of long words (words with 7 or more letters) per chapter. The horizontal dotted line is the the 0.95 quantile of the distribution of M 3.037, in absolute value.

Chapter 10

Supplementary remarks

We now mention some further relevant research topics that was not prioritized in this thesis.

More than one change-point

In some applications, such as our application on the novel *Tirant lo Blanch*, we have good reason to expect that there is one, and only one, change-point. In other cases we might have good reason to assume that there are more than one change-point. For British mining disasters we found two likely change-points. If we have a prior belief that there are two change-points τ_1 and τ_2 , we might divide the sequence of observations into three subsets, where $Y_1 = y_1, \dots, y_{\tau_1}$, $Y_2 = y_{\tau_1+1}, \dots, y_{\tau_2}$ and $Y_3 = y_{\tau_2+1}, \dots, y_n$, for $\tau_1 < \tau_2$, and test the homogeneity of Y_1, Y_2 and Y_3 separately. Letting $M(Y_i)$ be the M-test for the subsequence Y_i and applying the M-test on each of Y_1, Y_2 and Y_3 , we get a level α confidence set by testing each subset at level $\alpha^{1/3}$. We can find a confidence set at level α for τ_1 and τ_2 by

$$R(\alpha) = \{\tau_1, \tau_2 : M(Y_1), M(Y_2) \text{ and } M(Y_3) \text{ accepted at level } \alpha^{1/3}\}$$

Dependent observations

We have only looked at theory for the case of independent observations. In many practical cases observations are not independent.

Testing more than one parameter

We have focused on testing the homogeneity of one parameter at the time. For $y_i \sim f(y_i, \theta_1)$ where θ has dimension p , $\theta = (\theta_1, \dots, \theta_p)$, we have only tested a single θ_i at the time. Tests can also be constructed for more than one parameter at the time, see [Koning and Hjort, 2002](#).

Regression

Often in analysis, observation are dependent on covariates. Say y_i depends linearly on k covariates $x_1, \dots, x_k$, we then have a regression model

$$y_i = \psi_0 + \sum_{j=1}^k \psi_j x_j + \epsilon_i$$

where $\psi_0, \dots, \psi_k$ are parameters. The parameter ψ_1 might for example dependent on the value of the covariate x_1 , it might take two different values depending on whether x_1 is smaller or larger than some given constant. Then there is a change-point. A possible application is how the number of cigarettes a person smokes affects their risk of cancer. Let the observation y_i be the level of carboxyhaemoglobin (COHb) in a persons blood, and let their reported number of cigarettes smoked per day be a covariate. For people who smoke between 0 and 20 cigarettes per day, there is an approximate linear relationship between COHb and cigarette consumption. Above 20 cigarettes per day, however, the increase in COHb is proportionately smaller [Law et al., 1997](#).

Time series

For a time series $y_1, \dots, y_n$ where the observations are taken at equally spaced points in time there might for example be autocorrelation, then the correlation between two observations y_i and y_j is a function of the time between them. For a value δ , where $0 \leq \delta \leq 1$, the correlation between y_i and y_j could for example be

$$\text{Corr}(y_i, y_j) = \pm \delta^{|i-j|}$$

If the output variable depends linearly on its own previous values, we have an autoregressive model of order p such that

$$y_i = \phi_0 + \sum_{j=1}^p \phi_j y_{i-j} + \epsilon_i$$

where ϵ_i is normally distributed with expectation 0 and $\phi_1, \dots, \phi_p$ are parameters. We can look for change-points in the parameters.

Bayesian change-point model

One can also do a Bayesian approach to hypothesis testing.

Appendix A

Some R code

Test statistic

For a given number of observations n , we draw values of the statistic M_n , of standard normal observations, by

```
Hn_mu = function(N, tau, muL, muR, sig2L, sig2R){
  return( (muL-muR)/sqrt(sig2L/tau + sig2R/(N-tau)) )
}
Mn_mu <- function(myData){
  n = length(myData)
  testIndex = ceiling(.1*n):floor(.9*n)
  HnVec = rep(NA, length(testIndex))
  eps = testIndex[1]-1
  for (i in testIndex) {
    data_L = myData[1:i]
    data_R = myData[(i+1):n]
    mu_L = mean(data_L)
    mu_R = mean(data_R)
    sigma2L = var(data_L)
    sigma2R = var(data_R)
    HnVec[(i-eps)] = Hn_mu(n, i, mu_L, mu_R, sigma2L, sigma2R)
  }
  return(max(abs(HnVec)))
}
```

We find the limit distribution of the statistic M_n by

```
A <- function(x,s){
  n <- length(x)
  tau <- floor(n*s)
  x.new <- x[1:tau]
  a <- (1/sqrt(n))*sum(x.new)
  return(a)
}
q <- function(x,s){
  return(
    (abs(A(x,s) - s*A(x,1))) / sqrt(s*(1-s))
  )
}
M <- function(x){
  n <- length(x)
  c <- ceiling(0.1*n)
  d <- floor(0.9*n)
  vec <- rep(0,n)
  for (i in c:d) {
    vec[i] <- q(x,i/n)
  }
  return(max(abs(vec)))
}
## Draw values from M
B = 100000
M_vec <- rep(NA, B)
for (i in 1:B){
  M_vec[i] <- M(rnorm(10000,0,1))
}
```

```

}
## find percentiles, here 0.99
Mdata_sorted = sort(M_vec)
Mdata_sorted[B*0.99]

```

Confidence curves

The confidence curve based on the M-test. Here we have the code for British mining disasters, assuming a Poisson distribution.

```

Hn = function(n, tau, lambda1, lambda2){
  return( (lambda1-lambda2)/sqrt(lambda1/tau + lambda2/(n-tau) ) )
}
Mn <- function(myData){
  n = length(myData)
  eps = floor(0.1*n)
  testIndex = (eps+1):(n-eps)
  HnVec = rep(NA, length(testIndex))
  for (i in testIndex) {
    data_L = myData[1:i]
    data_R = myData[(i+1):n]
    lambda_L = mean(data_L)
    lambda_R = mean(data_R)
    HnVec[(i-eps)] = Hn(n, i, lambda_L, lambda_R)
  }
  return(max(abs(HnVec)))
}
cc_M = function(dataObs){
  N = length(dataObs)
  eps = floor(0.1*N)
  index = (eps+1):(N-eps)
  R = rep(NA,N)
  for (tau in index) {
    dataL = dataObs[1:tau]
    dataR = dataObs[(tau+1):N]
    a = Mn(dataL)
    b = Mn(dataR)
    r = max(a,b)
    root.alpha = max(which( (Mdata_sorted_FINAL <= r)==TRUE ) )
                                     /length(Mdata_sorted_FINAL)
    R[tau] = root.alpha*root.alpha
  }
  R = na.omit(R)
  return(cbind(index, R))
}

```

The confidence curve based on profile log-likelihood, again the code for British mining disasters, assuming a Poisson distribution.

```

# log-likelihood function:
l_prof_tau <- function(data, tau) {
  dL = data[1:tau]
  dR = data[(tau+1):N]
  lambda_L = mean(dL)
  lambda_R = mean(dR)
  return(sum(log(dpois(dL, lambda_L))) + sum(log(dpois(dR, lambda_R))) )
}

#Find maximum log-likelihood.
max_l = function(data){
  likelihood = rep(NA, N)
  for (i in eps:(N-eps)) {likelihood[i] = l_prof_tau(data, i)}
  tau.ML <- which.max(likelihood)
  return(tau.ML)
}
tau.ML = max_l(accidents)

# Deviance
D <- function(data, tau, tau_ML){
  return( 2*(l_prof_tau(data, tau_ML) - l_prof_tau(data, tau)) )
}

```

```

}

# Confidence curve
B <- 10000
ccL <- rep(NA, N)
for (tau in (eps+1):(N-eps)) {
  lambdaL.star = mean(accidents[1:tau])
  lambdaR.star = mean(accidents[(tau+1):N])
  I_vec = rep(NA, B)
  for (i in 1:B) {
    y_star_L = rpois(tau, lambdaL.star)
    y_star_R = rpois((N-tau), lambdaR.star)
    y_star = c(y_star_L, y_star_R)
    tau_ML.star = max_l(y_star)
    I_vec[i] = as.integer(D(y_star, tau, tau_ML.star) < D(accidents, tau, tau_ML))
  }
  ccL[tau] <- mean(I_vec)
}

```


Bibliography

- [Billingsley, 1968] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley.
- [Casella and Berger, 2002] Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Brooks/cole, Cengage Learning.
- [Clauset, 2018] Clauset, A. (2018). Trends and fluctuations in the severity of interstate wars.
- [Cunen et al., 2018] Cunen, C., Hermansen, G., and Hjort, N. L. (2018). Confidence distributions for change-points and regime shifts. *Journal of Statistical Planning and Inference*.
- [Cunen et al., 2020] Cunen, C., Hjort, N., and Nygård, H. (2020). Statistical sightings of better angels: Analysing the distribution of battle-deaths in interstate conflict over time. *Journal of Peace Research*, 57.
- [Girón et al., 2005] Girón, J., Ginebra, J., and Riba, A. (2005). Bayesian analysis of a multinomial sequence and homogeneity of literary style. *The American Statistician*, 59(1):19–30.
- [Huber, 1981] Huber, P. J. (1981). *Robust Statistics*. Wiley-Interscience.
- [Jarrett, 1979] Jarrett, R. G. (1979). A note on the intervals between coal-mining disasters. *Biometrika*, 66(1):191–193.
- [Koning and Hjort, 2002] Koning, A. and Hjort, N. L. (2002). Tests for constancy of model parameters over time. *Journal of nonparametric Statistics*.
- [Law et al., 1997] Law, M., Morris, J., Watt, H., and Wald, N. (1997). The dose-response relationship between cigarette consumption, biochemical markers and risk of lung cancer.
- [Moreno et al., 2005] Moreno, E., Casella, G., and Garcia-Ferrer, A. (2005). An objective bayesian analysis of the change point problem. *Stochastic Environmental Research and Risk Assessment*, 19:191–204.
- [Ross, 1983] Ross, S. M. (1983). *Stochastic processes*. Jhon Wiley & Sons.
- [Sarkees and Wayman, 2010] Sarkees, M. R. and Wayman, F. (2010). Resort to war: 1816 - 2007.
- [Vaart, 1998] Vaart, A. W. v. d. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.