Testing for a Change in Mean of a Weakly Stationary Time Series

Thesis to be submitted in partial fulfillment of the requirements for the degree

of

Integrated Master of Science in Mathematics and Computing

by

Shreyas Kowshik
17MA20039

Under the guidance of

Prof. Buddhananda Banerjee

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

Department of Mathematics
Indian Institute of Technology,
Kharagpur
India - 721302

CERTIFICATE

This is to certify that we have examined the thesis entitled **Testing for a Change in Mean of a Weakly Stationary Time Series**, submitted by the student **Shreyas Kowshik**(Roll Number: *17MA20039*) a graduate student of **Department of Mathematics** in partial fulfillment for the award of degree of Integrated Master of Science in Mathematics and Computing. We hereby accord our approval of it as a study carried out and presented in a manner required for its acceptance in partial fulfillment for the Post Graduate Degree for which it has been submitted. The thesis has fulfilled all the requirements as per the regulations of the Institute and has reached the standard needed for submission.

Buddhananda Banerjee

Department of Mathematics
Indian Institute of Technology,
Kharagpur

Place: Kharagpur
Date: 19/04/2022

ACKNOWLEDGEMENTS

For the work done in this thesis, I would like to thank Prof. Buddhananda Banerjee and Prof. Radhendushka Srivastava (Department of Mathematics, IIT Bombay) for all their guidance and support throughout the project. Most of the ideas in this work are a result of the weekly brainstorming meetings with them. They were the ones who introduced the basic form of the proposed statistic in this work and guided me throughout the semester. I am also thankful to my family and friends who constantly supported me throughout the semester in such difficult times. Without them this work would not have been possible.

Shreyas Kowshik
IIT Kharagpur

Date: 19/04/2022

ABSTRACT

Change-point detection in a time series is usually dealt with as a testing problem to examine its existence at an intermediate point of a given sequence. Specifically, we consider the detection of a mean-shift as a test for change in parameter. Standard approaches for this involve obtaining a consistent estimate for the long-run variance which is a nuisance parameter. This estimate requires selecting a bandwidth parameter that can be data-dependent. However, under the presence of a change-point, the bandwidth estimate can be severely biased and can lead to non-monotonic power on deviation from the null hypothesis. To circumvent this, self-normalization based test-statistics have been introduced which have monotonic power under the alternative hypothesis. In this work, we propose a statistic that has a sharper power rise on deviation from the null hypothesis compared to the existing ones. We evaluate its performance with exact simulation and theoretically analyze its features. Finally, we discuss approaches to estimate the variance term involved in the statistic and show simulation studies on the same. By obtaining sharper power rise for our proposed statistic, we validate the improvement of our proposed statistic in a general setting.

Keywords: Change-Point, Mean-shift, Self-normalization, long-run variance.

Contents

List of Figures

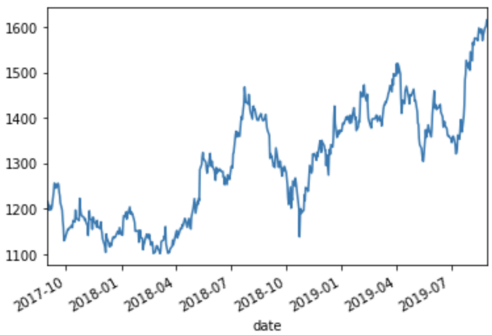
Chapter 1

Introduction

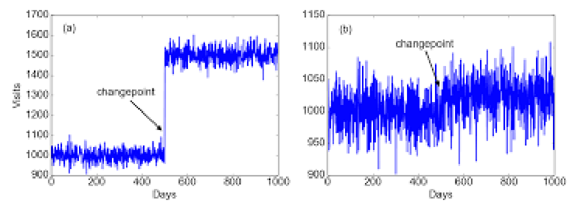
A Time-Series is a collection of random variables $\{X_t | t \in T\}$ over a time index set T, which might be a finite, countably infinite or an uncountable set. What we observe in reality are realizations of these random variables i.e. the data set is $\{X_1 = x_1, \cdots, X_n = x_n\}$, where the x_is are some numeric or categorical values. A lot of real-world data can be naturally modeled as a time-series. For example : The population of India, Price of a stock, Number of daily visitors to a museum, etc. Figure 1.1a shows an example of a time-series : a stock price varying over time.

The mean of a time series $\{X_t\}$ is defined as : $\mu(t) = \mathbb{E}\left[X_t\right]$. The autocovariance function is defined as : $\gamma_X(r, s) = \text{Cov}(X_r, X_s) = \mathbb{E}\left[\big(X_r - \mu_X(r)\big)\big(X_s - \mu_X(s)\big)\right]$. If a time-series follows some specific mathematical properties, analysis and forecasting it becomes easier. Time-Series analysis involves using such properties to develop tests and estimates of the parameters involved and analysing them in such settings.

A change-point is said to occur in a time-series if the distribution of the series changes abruptly at some time-step. For instance, during a week, the price of a stock may abruptly change its mean, there may be a security attack on a networked system affecting its statistics, etc. The change can be for instance in the mean, the variance, etc. Detection of such structural changes is extremely important in risk-sensitive applications. For instance, if an automated-trading-system uses some time-series model to forecast future prices and decide what trades to execute, a change in distribution can severely effect its predictions and lead to massive losses for the firm. If the output quality of a production process changes sharply due to a defect and remains undetected, all following samples will get wasted. Thus detecting if a change has occurred in the series becomes extremely important. Figure 1.1b shows



(a) Time Series Example : Stock Price data



(b) Change in Mean in a Time Series

Figure 1.1: Time Series and Changepoint Example

an example of a change in mean occurring in a time series. The figure on the left is very extreme in terms of the magnitude of change. However the change in mean is much more subtle in the figure on the right. One thus needs to develop statistical tests to detect such subtle changes reliably.

There exists a lot of literature on applying change-point detection in real-world settings. Aue et al. (2012) study change-point detection in the context of capital-asset-pricing of a portfolio. They test whether the β ratios of financial assets are stable overtime before using them in their portfolio optimization model. Chu et al. (1996) develop tests for high-frequency data such as streaming stock prices in high-frequency trading. Page (1954) develop change-point testing for detecting changes in the quality of an output of a continuous production process.

In the next section we highlight the problem statement and describe some approaches for detecting the change in the mean of a time series.

1.1 Literature Review

By virtue of tradition, an existence of change-point has been formulated as a testing of hypothesis problem by Aue & Horváth (2013). The null hypothesis denotes the absence of any change and the alternative denotes at least one change occurs in the distribution of the series at an intermediate point. Under the assumption of i.i.d. samples, Csorgo & Horváth (1997) and Darkhovski (1994) provide various methods for testing for a change-point. However, these ideas are easily not extendable when there is temporal dependence in the series. Suitable modifications are needed to make them work as can be seen from Tang & MacNeill (1993).

While testing for a change in mean in a weakly stationary time series, one needs to estimate the long-run variance of the series. This involves a kernel based estimate and requires selection of a bandwidth parameter. Andrews (1991) lists different kernels for estimating the long-run variance and methods to properly select the bandwidth of such kernels. This gives a data-dependent bandwidth. However, under the alternative hypothesis, Shao & Zhang (2010) argue that this can lead to the non-monotonic power problem. In other words, as one deviates from the null-hypothesis further away, the power of the test keeps on decreasing. This is highly undesirable foe a testing methodology. To overcome this, Shao & Zhang (2010) propose a self-normalized test statistic, extending the self-normalization idea from Lobato (2001) and Shao (2010). However, they show a naive extension is still an issue and thus derive a new test-statistic. This new statistic takes into consideration the presence of a change-point while doing the normalization. Moreover, it also generalises the test to cases of change in median and spectrum.

The question that we thus pose is : Is there a better statistic to test for a mean shift, in terms of having a sharper power rise on deviation from the null hypothesis ? We propose a better statistic and compare it with the statistic provided by Shao & Zhang (2010) and further theoretical and simulation analysis of this statistic is done.

Chapter 2

Changepoint detection in Time Series

2.1 Problem Statement and Notation

Change-point detection is a problem of identification of the existence of an abrupt change in the distribution of a time series. This can be formulated as a hypothesis testing problem. The null hypothesis specifies that there is no change. The alternative hypothesis can be framed in multiple ways. For instance it can specify there is only one change. It can also specify that there are multiple change-points in the series. Specifically, for the purpose of this work, we concentrate on testing for a change in mean of a univariate series with only one change-point. Formally, given a random sample $X_1 \cdots X_n$ from a weakly stationary time series $\{X_t\}$, we are interested in testing the following :

$$H_0 : \mathbb{E}[X_1] = \cdots = \mathbb{E}[X_n]$$

$$\text{vs}$$

$$H_1 : \mathbb{E}[X_1] = \cdots = \mathbb{E}[X_{k^*}] \neq \mathbb{E}[X_{k^*+1}] = \cdots = \mathbb{E}[X_n]$$

where, $1 \leq k^* < n$ is unknown.

A class of statistics to detect change in mean is built upon the cumulative-sum (CUSUM) process. Defining $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ the CUSUM process is defined for $\{X_t\}$ as :

$$T_n([nt]) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nt]} (X_t - \overline{X}_n) \tag{2.1}$$

Under appropriate moment and weak dependence assumptions as in Phillips (1987), the following holds under the null hypothesis :

$$T_n(\lfloor nt \rfloor) \implies \sigma(B(t) - tB(1)) \tag{2.2}$$

where, ' $\implies$ ' denotes weak convrgence in $D[0,1]$ endowed with the Skorokhod topology, $\sigma^2 = \lim_{n\to\infty} n\text{Var}(\overline{X}_n) = \sum_{h\in\mathbb{Z}} \gamma(h)$ is the long run variance and $B(r)$ is the one-dimensional standard Brownian motion on $[0,1]$ and $k \in [0,1]$. This defines the Kolmogorov-Smirnov Test Statistic as :

$$KS_n = \sup_{t\in[0,1]} \left| \frac{T_n(\lfloor nt \rfloor)}{\hat{\sigma}_n} \right| \tag{2.3}$$

where $\hat{\sigma}_n$ is a consistent estimator of σ.

Some kernel based methods are commonly used to estimate σ^2 by selecting a data-dependent bandwidth parameter. The estimates of the bandwidth can be heavily biased under the alternative hypothesis. This may lead to non-monotonic power variation upon deviation from the null hypothesis; see Shao & Zhang (2010).

2.2 Self Normalizing Statistics

Self normalization in simple words is the process of normalizing a statistic with its own point-wise standard deviation. A common example of a self-normalized process is the student-t statistic. Shao (2010) highlight using self-normalization in the context of time series data. This idea is then extended by Shao & Zhang (2010) to develop a statistic for detecting a change in mean.

Thus, to avoid selection of a bandwidth parameter in the estimation of $\hat{\sigma}_n$, one can use a self-normalizing procedure to compute the statistic :

$$\widetilde{KS}_n = \sup_{t\in[0,1]} \left| \frac{T_n(\lfloor nt \rfloor)}{D_n} \right| \tag{2.4}$$

where $D_n^2 = n^{-2}\sum_{t=1}^{n}(\sum_{j=1}^{t}(X_j - \bar{X}_n)^2)$. This avoids direct estimation of σ^2.

This statistic however, still suffers from the non-monotonic power problem. This is due to its normalizer i.e. D_n not incorporating the change-point alternative. To overcome this, the following statistic was proposed in Shao & Zhang (2010) :

$$S_{t_1,t_2} = \sum_{j=t_1}^{t_2} X_j \text{ if } t_1 \leq t_2, 0 \text{ otherwise} \tag{2.5}$$

$$V_n(k) = n^{-2}[\sum_{t=1}^{k}(S_{1,t} - (t/k)S_{1,k})^2 + \sum_{t=k+1}^{n}(S_{t,n} - (n-t-1)/(n-k)S_{k+1,n})^2] \tag{2.6}$$

$$G_n = \sup_{k=1,\cdots,n-1} T_n(k)V_n^{-1}(k)T_n(k) \tag{2.7}$$

Under a set of moment and weak dependence assumptions as highlighted in Phillips (1987), it can be shown that :

$$G_n \xrightarrow{D} \sup_{r\in[0,1]} (B(r) - rB(1))'V^{-1}(r)(B(r) - rB(1))$$

$$V(r) = \int_0^r (B(s) - (s/r)B(r))^2 ds + \int_r^1 (B(1) - B(s) - (1-s)/(1-r)(B(1) - B(r)))^2 ds$$

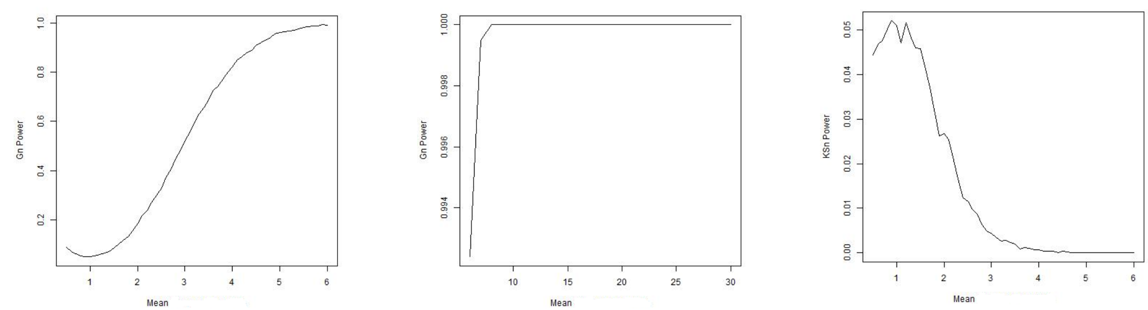
The key property of this statistic is that the denominator varies according to k which was not the case with $\widetilde{KS}_n$. Thus this statistic normalizes each location differently which is beneficial when a change-point is actually present. We verified in our simulations that this does not suffer from the non-monotonic power problem like $\widetilde{KS}_n$.

2.3 Simulations

For verifying the power of the test as one deviates from the null hypothesis, monte carlo simulations were run. The statistics were first simulated under the null hypothesis for 10000 iterations and cutoff values for the asymptotic distributions were obtained at 95% levels. These were then used to obtain the power for each of the statistics. We assumed the following data generating $AR(1)$ process :

$$X_t = \begin{cases} 1 + \rho s_{t-1} + \epsilon_t, & \text{if } 1 \leq t \leq N/2 \\ \eta + \rho s_{t-1} + \epsilon_t, & \text{if } N/2 \leq t \leq N \end{cases}$$

where $\epsilon_t \sim N(0,1)$.



(a) G_n power;$\eta \leq 5$ (b) G_n power;$5 \leq \eta \leq 30$ (c) $\widetilde{KS}_n$ power;$\eta \leq 5$

Figure 2.1: Power curves for $G_n, \widetilde{KS}_n$. X-axis : η, Y-axis : Power of test

We vary η in the range $[0, 30]$. $\eta = 1$ denotes the null hypothesis and the other values denote a deviation from the null hypothesis. $N = 200, \rho = 0.7$ were used for the sample size and the $AR(1)$ coefficient. The results of this simulation are shown in Figure 2.1.

It is clearly evident that the power of the $\widetilde{KS}_n$ statistic decreases as the value of η moves away from $\eta = 1$ which is highly undesirable. On the other hand, G_n has a monotonic power curve which increases to 1.0. As one moves further away from the null hypothesis, the tendency to reject the null hypothesis also increases in this case.

Chapter 3

Proposed Self-Normalizing Test-Statistic

3.1 Introduction

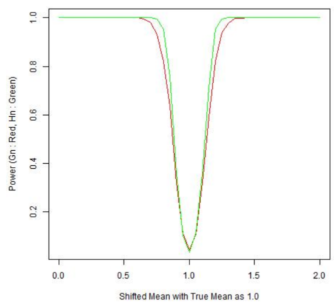
We propose a new test statistic that empirically outperforms the G_n statistic as used previously. First, the basic form of the statistic is introduced after which we look at some theoretical characteristics of the statistic and refine it further.

Let $\mathbf{X} = \{X_1, \cdots, X_n\}$ be n samples from a time series. Then, the proposal takes the following form :

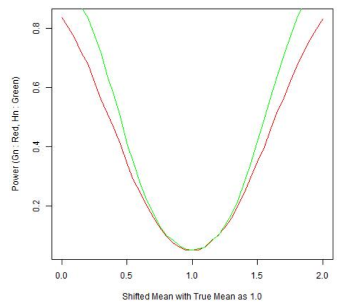
$$T_r(\mathbf{X}) = \frac{1}{n}\left[\left(1 - \frac{r}{n}\right)\sum_{i=1}^{r} X_i + \left(-\frac{r}{n}\right)\sum_{i=r+1}^{n} X_i\right]$$

$$H_n = \sup_{r=1,\cdots,n-1} \frac{\sqrt{n}T_r(\mathbf{X})}{\sqrt{\sum_{|h|<n} w(r,h,n)\gamma(h)}}$$

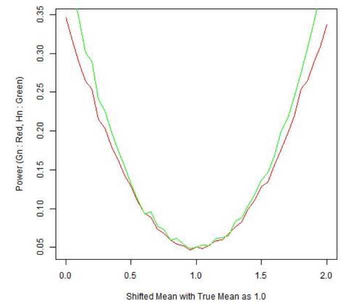
where $w(r, h, n)$ is a weighting function the details of which are available in section 3.4 and $r \in \{1, \cdots, n-1\}$. This statistic has a different normalizer compared to G_n, with the weighting function introduced as a consequence of the form of $Var(T_r(\mathbf{X}))$. We next show exact simulation studies of using this statistic on $AR(1)$ processes and analyse the power curves.



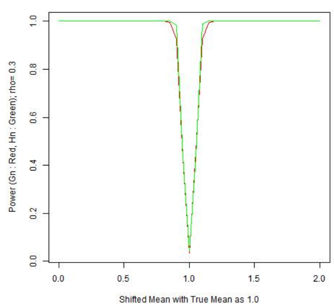
(a) $\rho = 0.7$, $\sigma_{WN} = 0.1, k = 100$



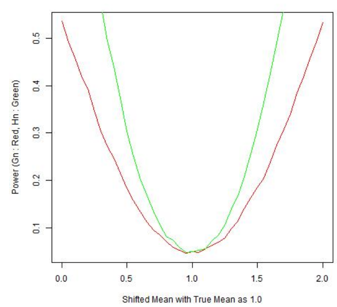
(b) $\rho = 0.7$, $\sigma_{WN} = 0.5, k = 100$



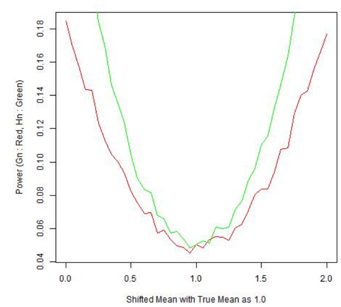
(c) $\rho = 0.7$, $\sigma_{WN} = 1, k = 100$



(d) $\rho = 0.3$, $\sigma_{WN} = 0.1, k = 100$



(e) $\rho = 0.7$, $\sigma_{WN} = 0.5, k = 50$



(f) $\rho = 0.7$, $\sigma_{WN} = 1, k = 50$

Figure 3.1: Power Curves comparing G_n v/s H_n statistics. Green : H_n, Red : G_n. On X-axis is plotted the new mean after the change-point with mean before as $\mu = 1$. On Y-axis is the power of the corresponding statistic's test.
ρ:$AR(1)$ coefficient, k:change-point location, $n = 200$, σ_{WN}:White-Noise Std Deviation

3.2 Simluation Studies

We simulate H_n under the null-hypothesis for 10000 iterations and obtain the cutoff values of the empirical distribution at 95% level. Then these are used to construct the power curve under the alternative hypothesis. As before, we keep the mean under the null hypothesis as $\mu = 1$ and vary under the alternative hypothesis as $0 \leq \mu \leq 2$. The process simulated is an $AR(1)$ process. A sample size of $n = 200$ is used and the change-point is varied as $k \in \{50, 100\}$. The variance parameter of the white-noise process is also varied as $\sigma_{WN} \in \{0.1, 0.5, 1\}$.

3.3 Experimental Analysis

The results of the simulation are shown in Figure 3.1. The power change for H_n is much sharper in general compared to G_n as one deviates from the null-hypothesis. This is particularly evident over a wide range of varying parameters. This affirms that there is good reason to investigate H_n and evaluate its performance more extensively.

Figure 3.2: Plot of $f(\alpha, \beta) = \frac{w(n\alpha, n\beta, n)}{n}$ as a function of $\alpha = \frac{r}{n}, \beta = \frac{l}{n}$ for lag l.

3.4 Theoretical Analysis

We now focus on the mathematical properties of the above statistic. The idea is to obtain a normalization of the statistic by dividing the numerator by its own variance. Observe that $\mathbb{E}[T_r(\mathbf{X})] = 0$. We obtain the variance of $T_r(\mathbf{X})$ as follows :

$$Var\left[\sqrt{n}T_r(\mathbf{X})\right] = \frac{1}{n}Var\left[\mathbf{a}^T\mathbf{X}\right] = \frac{1}{n}\mathbf{a}^T\Sigma\mathbf{a} \tag{3.1}$$

where $\mathbf{a}$ is a vector given by :

$$\mathbf{a} = \left[\underbrace{\left(1-\frac{r}{n}\right), \cdots, \left(1-\frac{r}{n}\right)}_{\text{r times}}, \underbrace{\left(-\frac{r}{n}\right), \cdots, \left(-\frac{r}{n}\right)}_{\text{n-r times}}\right]^T$$

and Σ is the covariance matrix of the vector $\mathbf{X}$

$$\Sigma = \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(n-2) \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ \gamma(n-1) & \dots & \dots & \gamma(0) \end{bmatrix}$$

Some simple calculations show that :

$$\mathbf{a}^T \Sigma \mathbf{a} = \gamma(0) \left(\sum_{i=1}^{n} a_i^2 \right) + 2 \sum_{l=1}^{n-1} \gamma(l) \left(\sum_{i=1}^{n-l} a_i a_{i+l} \right)$$

We define $\alpha = \frac{r}{n}$ and $\beta = \frac{l}{n}$ for a lag l and the coefficient of $\gamma(l)$ in the above summation as $w(r, l, n)$ for a fixed r. Then,

$$w(r, 0, n) = \sum_{i=1}^{n} a_i^2 = r \left(1 - \frac{r}{n} \right)^2 + (n - r) \left(\frac{r}{n} \right)^2 = n\alpha(1 - \alpha) \quad \text{[On Simplifying]}$$

Let $a = 1 - \alpha$ and $b = -\alpha$. These are the co-efficients of the vector $\mathbf{a}$. Then, $w(r, l, n)$ can be shown to be :

$$w(r, l, n) = \sum_{i=1}^{n-l} a_i a_{i+l} = \sum_{i=1}^{r} a^2 \mathbb{1}_{\{i+l \le r\}} + \sum_{i=1}^{r} ab \mathbb{1}_{\{r < i+l\}} + \sum_{i=r+1}^{n-l} b^2 \mathbb{1}_{\{i+l \le n\}}$$

where $\mathbb{1}$ is the indicator function that takes value 1 if its condition is true and 0 otherwise. The above equation can be further simplified into the following cases :

$$w(r, l, n) = \begin{cases} n \left[\alpha \left(1 - \alpha \right) - \beta \left(1 - \alpha + \alpha^2 \right) \right], & \text{if } \alpha \ge \beta \\ -n\beta\alpha^2, & \text{if } \alpha < \beta \end{cases}$$

Thus, the variance of the numerator can effectively be expressed as :

$$Var[\sqrt{n} T_r(\mathbf{X})] = \mathbf{a}^T \Sigma \mathbf{a} = \sum_{l=-(n-1)}^{n-1} \gamma(l) w(r, l, n) \tag{3.2}$$

where $w(r, l, n)$ is the coefficient of $\gamma(l)$ in the summation above.

A plot of $f(\alpha,\beta) = \frac{w(n\alpha,n\beta,n)}{n}$ is given in figure 3.2. It can be seen that for small values of β, as α varies in $[0,1]$, the value of f first increases and then decreases. This is the sort of behavior that will also be more practical. This is because one would generally use kernel based estimates for estimation of $\sum_l w(r,l,n)\gamma(l)$ and the cutoff lag is generally of the orders of $[n^{1/2}]$, $[n^{1/3}]$, etc. leading to large values of β being zeroed out in the estimate.

In the next chapter, we will validate Equation 3.2 empirically by simulating $\sqrt{n}T_r(\mathbf{X})$ and obtaining its sample variance.

Chapter 4

Normalizing Factor of Proposed Statistic

4.1 Introduction

The variance of the proposed statistic's numerator $\sqrt{n}T_r(\mathbf{X})$ is given by equation 3.2. It involves a term of $\sum_l w(r,l,n)\gamma(l)$. What is really of interest here is the weighted summation of $\gamma(h)$ rather than the individual values and we try to obtain an estimate of this.

If one has a knowledge of the generative process underlying the series, it can be possible to obtain the auto-covariances $\gamma(l)$ in closed form. However this is not practical as one never knows what the underlying process will look like beforehand. All one can thus do is to get an estimate $\hat{\gamma}(l)$ such that it is consistent. One naive way to estimate $\gamma(l)$ is to use the sample auto-covariances as :

$$\hat{\gamma}(l) = \frac{1}{n}\sum_{j=1}^{n-|l|}\left(X_j - \overline{X}_n\right)\left(X_{j+|l|} - \overline{X}_n\right) \tag{4.1}$$

for each lag l. However this leads to an inconsistent estimate of the long run variance $\sum_l \gamma(l)$. What is generally done in the literature is to smooth the estimates out using a kernel function :

$$\hat{\sigma}_n^2 = \sum_{k=-l_n}^{l_n}\hat{\gamma}(k)\mathbb{K}\left(\frac{k}{l_n}\right) \tag{4.2}$$

n	ρ =0.2	ρ =0.3	ρ =0.5	ρ =0.7	ρ =0.9	ρ =0.95
32	0.7	1.2	2.4	4.3	10.2	16.6
64	0.9	1.5	3	5.4	12.9	20.9
128	1.1	1.8	3.8	6.8	16.2	26.3
256	1.4	2.3	4.8	8.6	20.4	33.1
512	1.7	2.9	6	10.9	25.7	41.7
1024	2.1	3.7	7.6	13.7	32.4	52.6

(a) Bartlett Kernel Bandwidth Values

n	ρ =0.2	ρ =0.3	ρ =0.5	ρ =0.7	ρ =0.9	ρ =0.95
32	1.0	1.4	2.5	4.5	12.1	21.6
64	1.1	1.6	2.9	5.2	13.9	24.8
128	1.3	1.9	3.3	5.9	16.0	28.5
256	1.5	2.2	3.8	6.8	18.4	32.7
512	1.7	2.5	4.4	7.8	21.1	37.5
1024	2.0	2.9	5.0	9.0	24.2	43.1

(b) QS Kernel Bandwidth Values

Table 4.1: Bandwidth values for a $AR(1)$ process with parameter ρ for Bartlett and Quadratic-Spectral Kernels.

Truncated: $$k_{TR}(x) = \begin{cases} 1 & \text{for } |x| \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

Bartlett: $$k_{BT}(x) = \begin{cases} 1 - |x| & \text{for } |x| \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

Parzen: $$k_{PR}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \leqslant |x| \leqslant 1/2, \\ 2(1 - |x|)^3 & \text{for } 1/2 \leqslant |x| \leqslant 1, \\ 0 & \text{otherwise} \end{cases}$$

Tukey-Hanning: $$k_{TH}(x) = \begin{cases} (1 + \cos(\pi x))/2 & \text{for } |x| \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

Quadratic Spectral: $$k_{QS}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$$

Figure 4.1: Kernel Formulations from Andrews (1991)

where $\mathbb{K}(.)$ is a kernel function and $l = l_n$ is a bandwidth parameter. By appropriate choice of the bandwidth l_n and the kernel, one can get consistent estimates of the long run variance.

Different kernels exist in literature such as Bartlett, Truncated, etc. Figure 4.1 as given in Andrews (1991) provides a wide variety of kernels. There are also sophisticated techniques to automatically select a suitable bandwidth given the sampled data. Andrews (1991) provide one such approach for bandwidth selection . Table 4.1 provides the automatically selected bandwidth values for the Bartlett and Quadratic Spectral kernels that are asymptotically optimal. For the rest of our work, we will be using this approach for obtaining the kernel based estimates.

One limitation that is still not addressed in the long-run-variance estimation, specific to our problem setup, is that there is not any assumption made by the statistic about the presence of a change-point. To address this, we propose to use a transformed series for obtaining the sample auto-covariances.

4.2 Transformation of Series

Given a sample $\mathbf{X} = \{X_1, \cdots, X_n\}$, define the following for $r \in \{1, \cdots, n-1\}$

$$\overline{X}_r = \frac{1}{r}\sum_{i=1}^{r} X_i \text{ and } \overline{\overline{X}}_r = \frac{1}{n-r}\sum_{i=r+1}^{n} X_i$$

The series is then transformed as follows :

$$\begin{aligned} Z_1 &= X_1 - \overline{X}_r \\ &\vdots \\ Z_r &= X_r - \overline{X}_r \\ Z_{r+1} &= X_{r+1} - \overline{\overline{X}}_r \\ &\vdots \\ Z_n &= X_n - \overline{\overline{X}}_r \end{aligned}$$

The transformed series $\mathbf{Z} = \{Z_1, \cdots, Z_n\}$ is used for computing the autocovariance estimates $\hat{\gamma}(h)$ as defined in the previous section.

The above transformation is intuitive to understand. For each location r, the mean of the sample points before that location and after that location are used to center the series. If there is a change in mean in the series, the values of $\overline{X}_r$ and $\overline{\overline{X}}_r$ will differ significantly at the change-point. Thus, this series tries to incorporate the effect of the change-point in its computation. Without the use of this, the auto-covariance estimates are bound to perform poorly as one deviates further away from the null hypothesis.

The above transformation can also be written as a matrix multiplication.

$$\mathbf{Z} = Br\mathbf{X}$$

$$Br = \left(\begin{array}{c|c} I_r - \frac{1}{r}11^T \quad \cdots & \mathbf{0} \\ \hline \mathbf{0} & \cdots \quad I_{n-r} - \frac{1}{n-r}11^T \end{array}\right)$$

4.3 Variance Estimation using Transformed Series

We have

$$Cov(\mathbf{Z}) = Br\Sigma Br^T$$

Any consistent estimator of the covariance matrix of the transformed series will converge to the above result. Under the null hypothesis, we want the variance of $\sqrt{n}T_r(\mathbf{X})$ estimated using the transformed series to be the same as without using the transformation.

More formally, given $Cov(X) = \Sigma$, we want to get the same variance of the numerator using $Cov(Z)$. One way to do this can be to create a quadratic form using a random vector $x \in R^n$ such that its value matches the variance without transformation. Mathematically speaking

$$\begin{aligned} x^T Cov(\mathbf{Z})x - a^T\Sigma a &= 0 \\ x^T Br\Sigma Br^T x - a^T\Sigma a &= 0 \\ (x^T Br - a^T)\Sigma(Br^T x - a) &= 0 \end{aligned}$$

Using the fact that $Br = Br^T$ the above identity holds when

$$Brx = a \tag{4.3}$$

Observe that $Bra = 0$ and hence a lies in the null space of Br and thus $Brx = a$ has no solution. To circumvent this issue, we propose adding a regularizing factor to Br to ensure that the system becomes consistent. To do so, we use the help of the following lemma as proved in Farid (2011) Lemma 2.1.

Lemma 4.3.1 *A strictly diagonally dominant matrix is invertible.*

Thus, we define $\widetilde{Br} = Br + \lambda I$ where I is the identity matrix and $\lambda > 0, \lambda \in R$. Using lemma 4.3.1, we can see that for each row, to ensure diagonal dominance the condition looks like $|1 - \frac{1}{r} + \lambda| > \frac{r-1}{r}$. Thus there exists a $\lambda > 0$ such that $\widetilde{Br}$ is diagonally dominant. Thus we can say that the given system can be made consistent for a given $\lambda > 0$. Thus we conclude that $\widetilde{Br}x = a$ forms a consistent system.

However this changes the transformation from what was formulated. In order to get to the same variance estimate under the null hypothesis, we pose an optimization problem. Given a sample $\mathbf{X} = \{X_1, \cdots, X_n\}$, and a $r \in \{1, \cdots, n-1\}$, define M_x to be the sample covariance matrix using only $\mathbf{X}$ and M_z the sample covariance matrix after transforming the sample to a series $\mathbf{Z}$. Using $\widetilde{Br} = Br + \lambda I$ and fixing a constraint threshold ϵ, we have

$$\begin{aligned} \min_{\lambda} \quad & \left| a^T M_x a - (\widetilde{Br}x)^T M_z (\widetilde{Br}x) \right| \\ \text{s.t.} \quad & \left\| \widetilde{Br}x - a \right\|_2^2 < \epsilon \\ & \lambda > 0 \end{aligned}$$

This is a quadratic optimization problem in a scalar λ. Since it involves minimization of the absolute value of a quadratic in a scalar, it is guaranteed to have minimizer. Thus there exists a $\lambda^* \in R$ such that λ^* is the minimizer of the above optimization problem.

Chapter 5

Simulation Studies with Variance Estimation

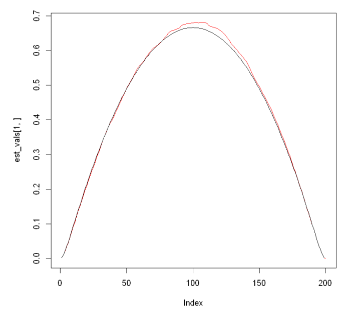
5.1 Experiments

We now describe in detail the experiments performed along with variance estimation. First, we simulate the closed form expression of equation 3.2 and compare it with the simulated variance of the numerator of H_n i.e. $\sqrt{n}T_r(\mathbf{X})$. The results of this simulation are given in Figure 5.2a. This was obtained by simulating $\sqrt{n}T_r(\mathbf{X})$ for 5000 iterations under the null-hypothesis for data from an $AR(1)$ process with $\rho = 0.7$. As can be clearly seen, both curves overlap significantly, thus affirming our hypothesis that equation 3.2 is correct.

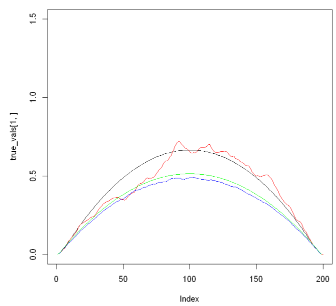
Next we try to estimate the variances using the transformed series. Simulations are run for 500 iterations for generating the above statistic values. Different kernels have been tried for the same : Quadratic Spectral Kernel and Bartlett Kernel and the bandwidth was manually selected for the given parameters. The λ value was obtained by doing a grid search in the range $[0.1, 3.0]$.

The results are shown in Figure 5.1. It is clearly evident that by tuning the kernel bandwidth parameter, we get variance estimates that are very close to the true value with the Bartlett kernel. However the value of the bandwidth becomes critical in order to get a good estimate. Even with using the Quadratic Spectral kernel, we are able to get estimates that are quite close to the true value, the difference being explained by the finite sample size and small number of iterations.

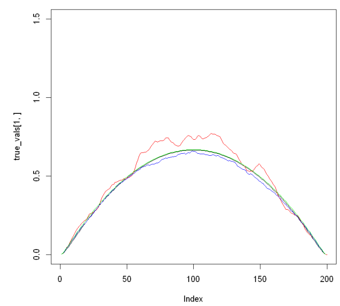
We now proceed to check the power curves obtained by using the above variance estimation procedure. Since the variance estimates are accurate, we expect the power



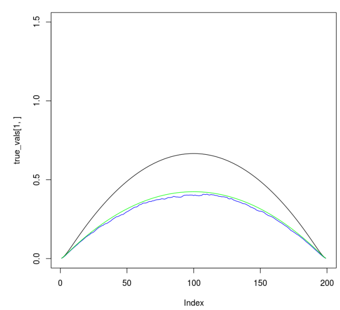
(a) True v/s Sample variance



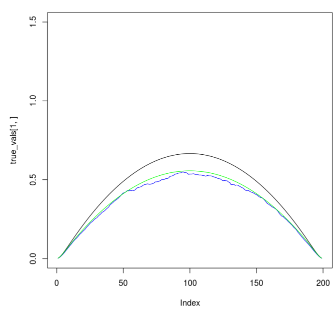
(b) Bartlett Kernel; Bandwidth=5.0 100 iterations



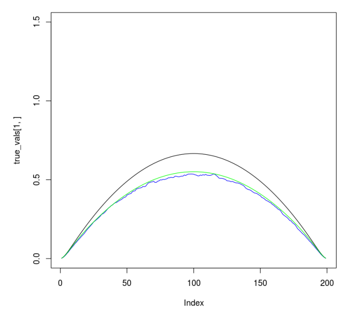
(c) Bartlett Kernel; Bandwidth=9.0 100 iterations



(d) QS Kernel; Bandwidth=5.0 100 iterations



(e) QS Kernel; Bandwidth=15.0 100 iterations

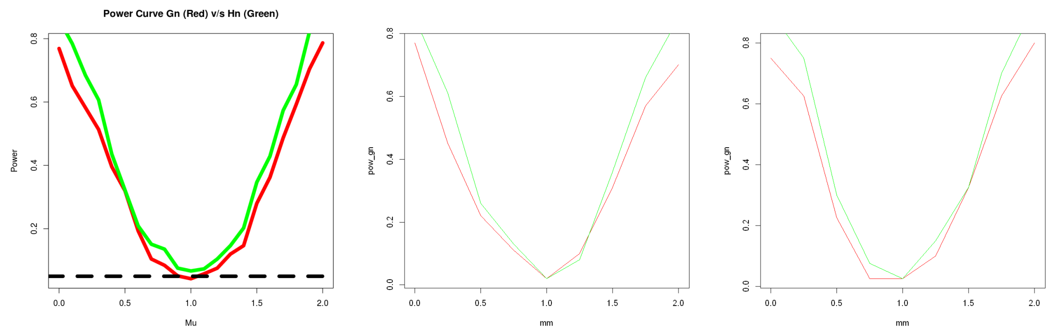


(f) QS Kernel; Bandwidth=30.0 100 iterations

Figure 5.1: Comparision of variances (X-axis : Index r, Y-axis : Variance value). Red : Sample variance of $\sqrt{n}T_r(\mathbf{X})$, Black : Variance predicted from equation 3.2, Blue : Variance predicted after using transformation with given kernel, Green : Variance predicted without using transformation with given kernel. $n = 200$

rise to be sharper for our proposed statistic H_n compared to G_n. The results are shown in Figure 5.2.

The null hypothesis is considered at a mean of $\mu = 1.0$ and the mean is varied in the range $[0.0, 2.0]$. Grid size for the mean variation is varied as $\{0.1, 0.25\}$. The number of iterations under the alternative hypothesis is mentioned in the figure label.



(a) Grid size=0.1
450 iterations

(b) Grid size=0.25
500 iterations

(c) Grid size=0.25
250 iterations

Figure 5.2: Power Curves (X-axis : Mean Value, Y-axis : Power), Red : Power of G_n, Green : Power of H_n $n = 200$

From Figure 5.2 it is clear that H_n outperforms G_n in terms of the power rise on deviation from the null hypothesis. This is checked across two different grid size values. The slightly discontinous nature of the curves is owing to the grid size value which can further be reduced and the number of iterations which are slightly on the lower side for these experiments. Given enough time and compute, these simulations can be scaled up to higher number of iterations and finer grids to get much sharper and continuous power curves as obtained in Figure 3.1.

Chapter 6

Conclusion and Future Work

We have surveyed and experimented with multiple test-statistics for mean-shift detection in a time-series. The KS_n statistic is a naive self-normalizing statistic that does not take into account the change-point which might be present under the alternative hypothesis. Consequently it suffers from non-monotonic power as we have highlighted. To overcome this, the G_n statistic was introduced which has monotonic power on deviation from the null hypothesis. We then proposed a new statistic that has a sharper power curve than the G_n statistic. Simulation studies were done to confirm this. Consequently the self-normalizer for this statistic was derived and was validated through simulations. We then proposed an approach to obtain a robust estimate of the variance term using a novel transformation to aid the given formulation. Using this estimate we ran simulations and validated our hypothesis that our proposed statistic H_n has a sharper power rise on deviation from null hypothesis on comparing it with the G_n statistic.

Future work would be to run simulations with varying types of time series such as ARMA processes and real world data. More work needs to be done on obtaining a closed from expression for the solution of the optimization problem posed in order to do variance estimation. This would reduce computation time by a huge margin and help scale up simulations. Another promising direction of future work would be to consider the given formulation in the Bayesian paradigm and see if there is any improvement in the overall performance.

Bibliography

Andrews, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica* **59**(3), 817–858.
URL: *http://www.jstor.org/stable/2938229*

Aue, A. & Horváth, L. (2013), 'Structural breaks in time series', *Journal of Time Series Analysis* **34**(1), 1–16.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9892.2012.00819.x*

Aue, A., Hörmann, S., Horváth, L., Hušková, M. & Steinebach, J. G. (2012), 'Sequential testing for the stability of high-frequency portfolio betas', *Econometric Theory* **28**(4), 804–837.
URL: *http://www.jstor.org/stable/23257658*

Chu, C.-S. J., Stinchcombe, M. & White, H. (1996), 'Monitoring structural change', *Econometrica* **64**(5), 1045–1065.
URL: *http://www.jstor.org/stable/2171955*

Csorgo, M. & Horváth, L. (1997), Limit theorems in change-point analysis.

Darkhovski, B. S. (1994), 'Nonparametric methods in change-point problems: A general approach and some concrete algorithms', *Lecture Notes-Monograph Series* **23**, 99–107.
URL: *http://www.jstor.org/stable/4355766*

Farid, F. (2011), 'Notes on matrices with diagonally dominant properties', *Linear Algebra and its Applications* **435**(11), 2793–2812.
URL: *https://www.sciencedirect.com/science/article/pii/S002437951100365X*

Lobato, I. N. (2001), 'Testing that a dependent process is uncorrelated', *Journal of the American Statistical Association* **96**(455), 1066–1076.
URL: *http://www.jstor.org/stable/2670252*

Page, E. S. (1954), 'Continuous inspection schemes', *Biometrika* **41**(1/2), 100–115.
URL: *http://www.jstor.org/stable/2333009*

Phillips, P. C. B. (1987), 'Time series regression with a unit root', *Econometrica* **55**(2), 277–301.
URL: *http://www.jstor.org/stable/1913237*

Shao, X. (2010), 'A self-normalized approach to confidence interval construction in time series', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**(3), 343–366.
URL: *https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2009.00737.x*

Shao, X. & Zhang, X. (2010), 'Testing for change points in time series', *Journal of the American Statistical Association* **105**(491), 1228–1240.
URL: *https://doi.org/10.1198/jasa.2010.tm10103*

Tang, S. M. & MacNeill, I. B. (1993), 'The effect of serial correlation on tests for parameter change at unknown time', *The Annals of Statistics* **21**(1), 552–575.
URL: *http://www.jstor.org/stable/3035607*