National Cheng Kung University

Mechanical Engineering

Stochastic Dynamic Data - Analysis and Processing

Hypothesis Test

Author:
Zhao Kai-Wen

Supervisor:
Chang Ren-Jung

November 29, 2012

Contents

1 Introduction

In the assignment, we do three knids of hypothesis

- Gaussian distribution hypothesis test
- Mean estimation hypothesis test
- Stationarity check hypothesis test

Generate Gaussian random signal, process it with Butterworth filter and test the hypothesis with Matlab primitives function.

2 Signal Processing System

2.1 Source Data Specs

Random signal is produced by matlab function with the specifications.

Function name	Mean Value	Standard Deviation	Data Length
normrnd()	0	1	100,000

2.2 Butterworth filter Specs

We use the Butterworth filter which is used is previous assignment

Order	Cutoff frequency	Settling Time
n = 2	1 rad/sec	5.96 sec

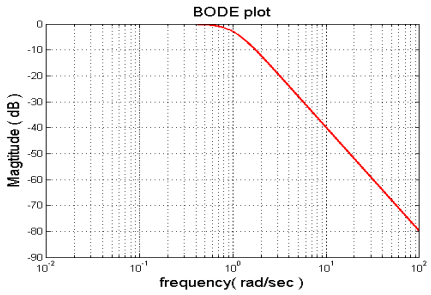

Figure 1: Bode Plot of Butterworth

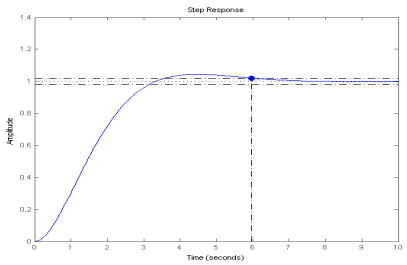

Figure 2: Settling Time

3 The Procedure of processing and test

3.1 Flow Chart

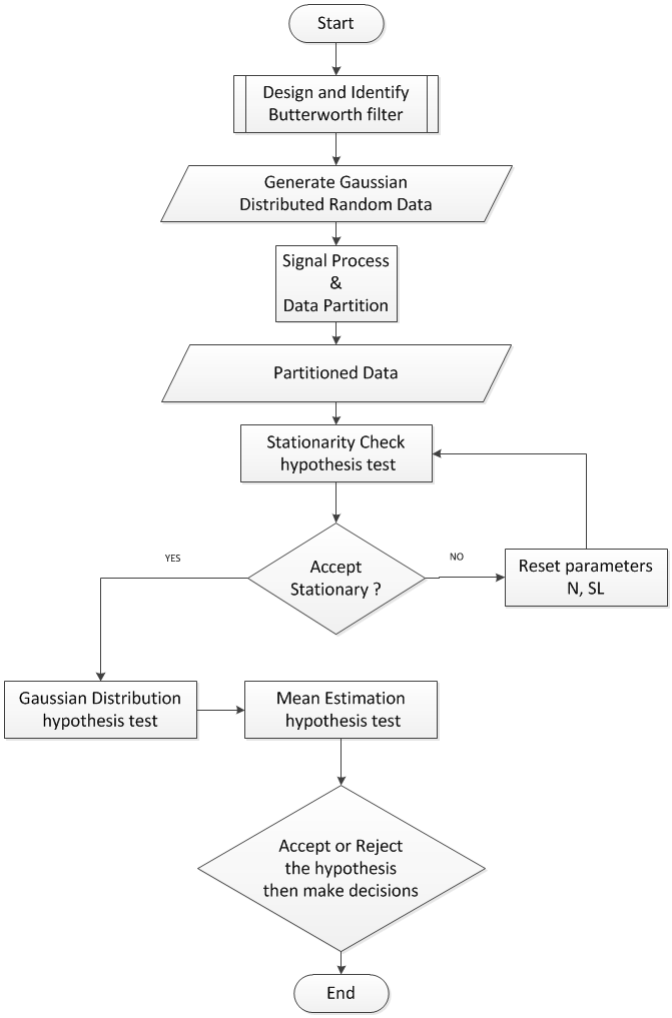

Figure 3: Processing Flow of the aissgnment

3.2 Data Partition

We processed the data with Butterworth filter and the specs of data are listed.

Method	Data Length	Sampling Rate
2^{th} Butterworth filter ($\omega_c = 1$)	100,000	10 Hz (time step = 0.1 sec)

Table 1: Specification of filtered Data

Consider that we want see more detailed change and influence when doing hypothesis test, we cut the data into many smaller sections. Because we could substitute ensemble data with time series in ergodic condition. We show how to partition the huge data into groups as the following graph.

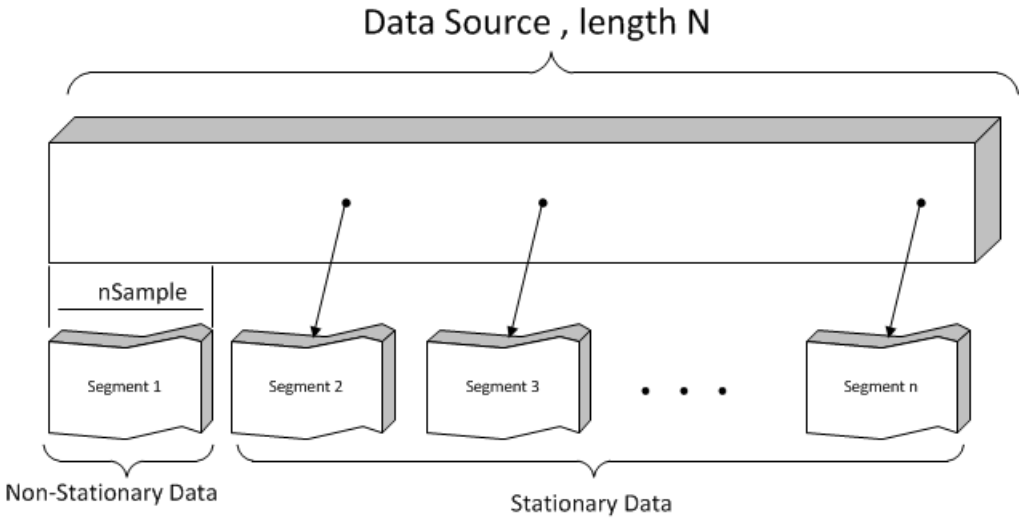

Figure 4: Data Partition and Storage

Based on what we learned before, the first one or two groups are nonstationary part. For the sake of checking them with hypothesis test, we do not get rid of them at first.

Number of Groups	Data Length of each Group
50	2,000

Table 2: Partitioned Data

4 Hypothesis Test

4.1 On Checking Stationarity

We download the function called checkstationary() to do the test first. We roughly predit with deterministic propertiy, settlint time, that 5.69 $sec \sim 60$ $data$ lies in the first section which contains 2000 data. So we do the test by each group.

Data Length	Observation Interval N	Siginificance Level	Method
1,000	100	$\alpha = 0.05$	criterion of series

Table 3: Parameters setting of checkstationary()

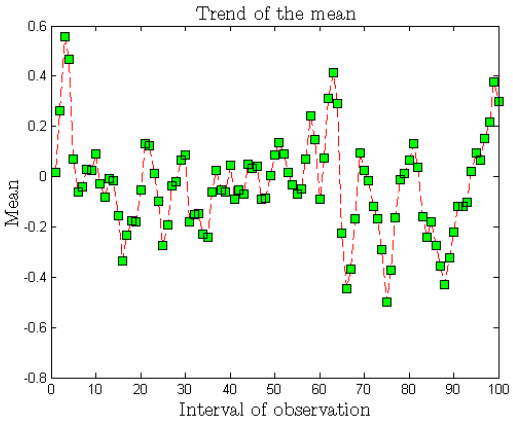

Figure 5: Trend of mean

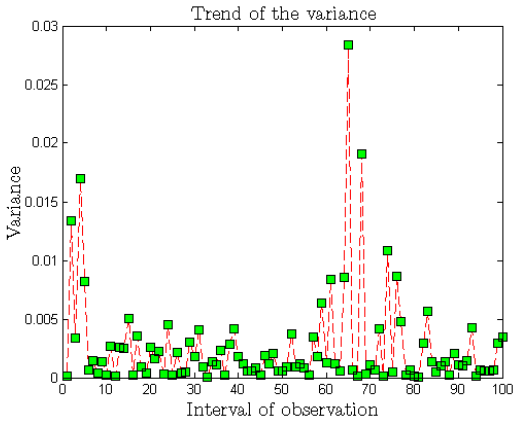

Figure 6: Trend of Variance

The mean value do not show much trend in the middle of section but it goes up on the end of data. On the right hand side, the variance trend is obvious non-stationary. It lies and sticks to the axis. Then we test block 2, it is stationary, as we gauss.

4.2 On Gaussian distribution

Now, we know that nonstationary part lies in the first group, so we eliminate the non-stationary section and do the test. Considre Jarque-Bera test is an asymptotic test that our samples in each group is at least larger than 2,000.

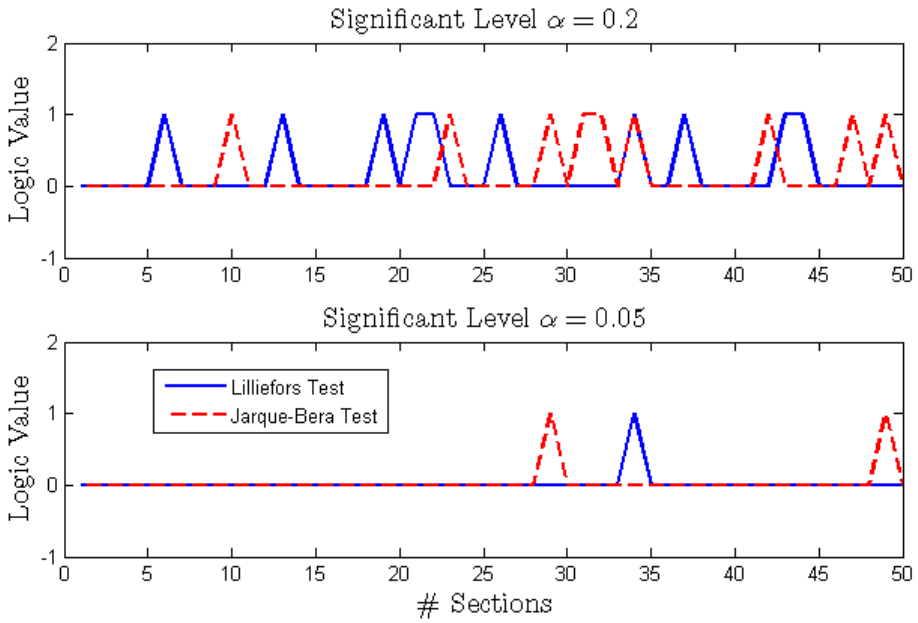

Figure 7: Gaussian Distributino Hypothesis

Significance Level	Type of Test	Reject	Accept
$\alpha = 0.2$	Lilliefors test	10	40
$\alpha = 0.2$	Jarque-Bera test	9	41
$\alpha = 0.05$	Lilliefors test	1	49
$\alpha = 0.05$	Jarque-Bera test	2	48

Table 4: Results of Gaussian Distributino Hypothesis

Intuitively, higher α indicates tough condition and cause more chance to reject. And we change the sample in group to greater number like 4,000. Jarque-Bera test rejects for less time than Lilliefors. Both results of tests tell us that we can trust the output signal is a normal distribution when $\alpha = 0.05$, but when $\alpha = 0.2$ the hypothesis could be further discussed .

4.3 On Mean Estimation

Apply z-test and student t-test to check that the mean value of the output signal. In ideal situation, the mean should be zero. We do the test and compare the resoult with different significance level and different data size seperately.

Parameters	Value	Condition
Significance Level α	0.05 / 0.10 / 0.25	at $N = 10,000$
Data Length N	10^3 / 10^4 / 10^5	at $\alpha = 0.05$

Table 5: Results of Gaussian Distributino Hypothesis

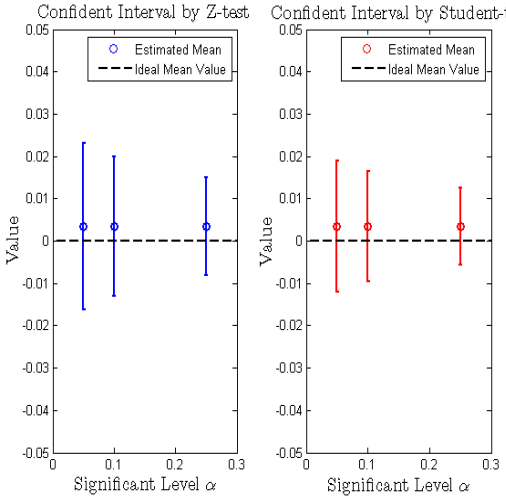

Figure 8: Confident Interval with α

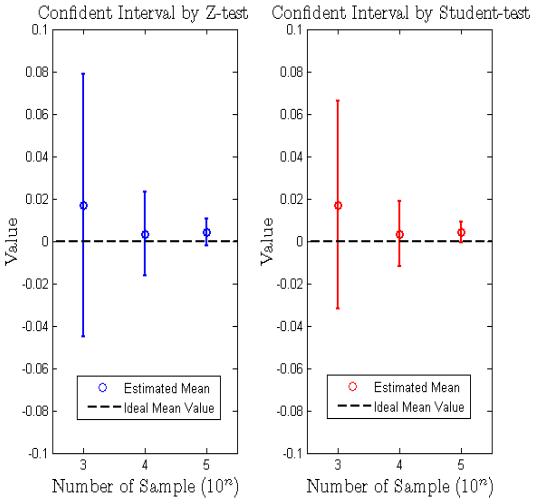

Figure 9: Confident Interval with N

We could easily get the conclusion :

- Significance Level increase, the interval narrows (slightly)
- Samples increase, the interval narrows (significantly)

According to the plot, the larger interval have more chance to cover the ideal value, can more data indicates closer to the ideal value. That is a kind of numerical verification of center limit theorem.
All in all, based on the hypothesis we do, we could accept that the output data are Gaussian at the significance level 5%.