National Cheng Kung University

Mechanical Engineering

Stochastic Dynamic Data - Analysis and Processing

Properties of Gaussian Process

Author:
Zhao Kai-Wen

Supervisor:
Chang Ren-Jung

November 8, 2012

Contents

1 Introduction

Gaussian process is an important method to model random procedure due to its formidable properties. Two of significant properties in random process are invariant and ergodic.

2 Invariant Property

2.1 Description

Gaussian process is quite prevalent in physical problems and often may be mathematically predictble. It possesses an important property that a Gaussian process undergoes a linear transformation, then the output will still be a Gaussian process.

2.2 Mathematical Verification

Suppose linear transformation from probability domain X to Y is designated T.

$$Y = T(X)$$

The density function and distribution of X domain are $f_X(x)$ and $F_X(x)$, so as Y domain. If our functions are strictly monotonic, the one-to-one correspondence between X and Y is

$$F_Y(y_0) = Prob\{Y \leq y_0\} = Prob\{X \leq x_0\} = F_X(x_0)$$

or represented in derivative form

$$f_Y(y) = f_X(x)|\frac{dx}{dy}|$$

Here we first verify the theorem with a **simple(static) case**.

$$Y = T(X) = aX + b$$

so that

$$X = T^{-1}(Y) = (Y - b)/a$$

$$dy/dx = 1/a$$

If X is assumed to be Gaussian with the density function

$$f_X(x) = \frac{1}{\sigma_x\sqrt{2\pi}} exp(\frac{-(x-\mu_x)^2}{2\sigma_x^2})$$

so we get

$$f_Y(y) = \frac{1}{\sigma_X\sqrt{2\pi}} exp(-\frac{\{(y-b)/(a\mu_X)\}^2}{2\sigma_x^2})|\frac{1}{a}|$$

which is the density function of another Guassian random variable having

$$\mu_Y = a\mu_X + b$$

and

$$\sigma_Y^2 = a^2\sigma_X^2$$

They still possess their Gaussian properties that we show a linear transformation of a Gaussian random variable produces another Gaussian random variable.

2.3 Numerical Verification

Another way, we use a first order LTI measurement system **(dynamic case)** to simulate and verify the invariant property.

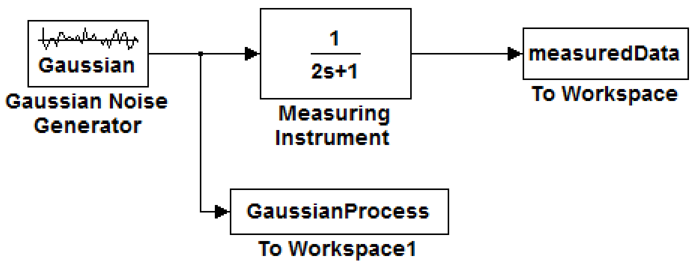


Figure 1: Measuring System

The system specs are listed

Measurement Parameters	Values
Samping rate	10 Hz
Time span	0 $\sim$ 1000 sec
Data length	10,000
System Bandwidth	0.5 rad/sec

Visualize the input and output data together, we could plot the scatter histogram. It is easier to observe the invariant. Notice that measuring would truncate data

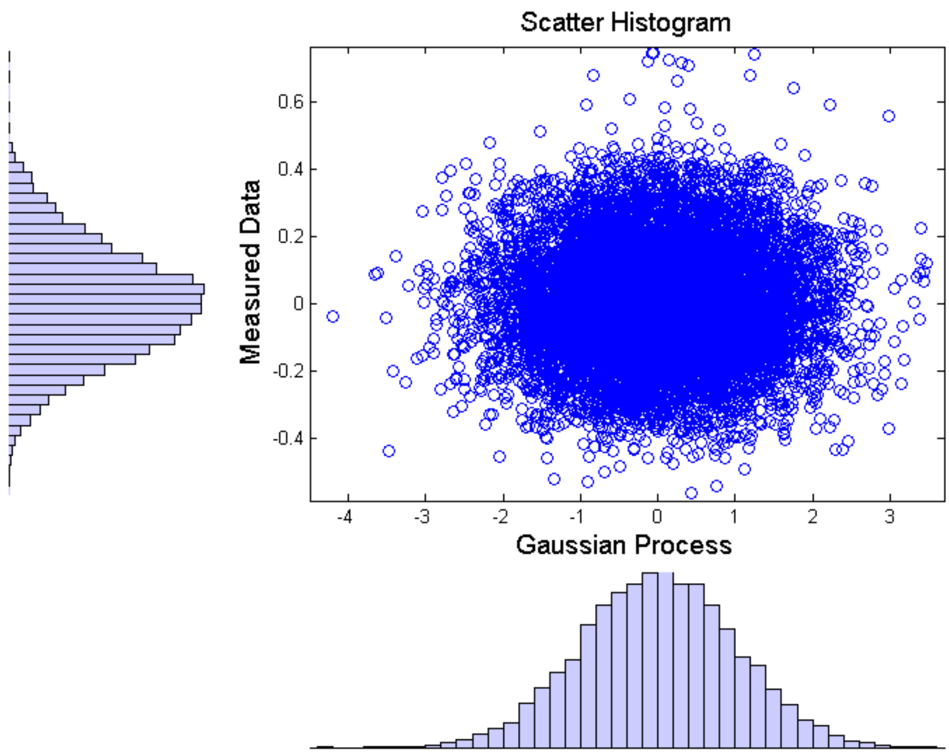


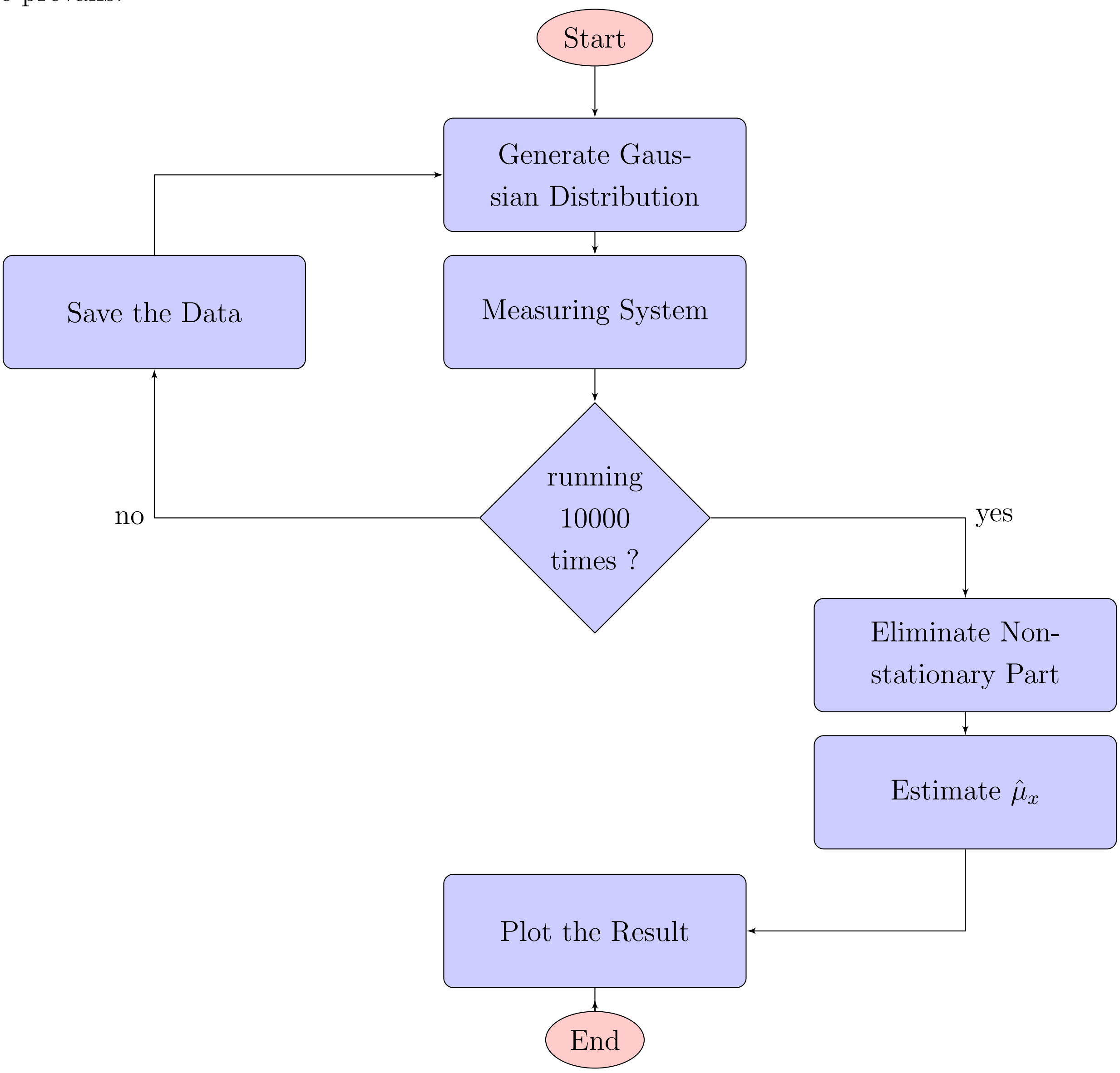
Figure 2: Scatter histogram

with high frequency, so scale of two axis are not equal.

3 Ergodic Property

3.1 Description

Ergodicity is a big picture in random data processing. In engineering view, we can calculate time averages on the arbitrary pair of sample funtions to replace ensemble average. I can not show with mathematical way, so numerical verification here prevails.



3.2 Numerical Verification

For weakly mathematically ergodic condition are

$$\mu(x,k) = \mu(x)$$

and

$$R_{xx}(\tau,k) = R_{xx}(\tau)$$

where k is the index of sample function. The autocorrelation function would be envelop decay. I would show these properties this section.

3.2.1 Measuring Instrument

Do the measurement system identification. With step response we could find settling time.

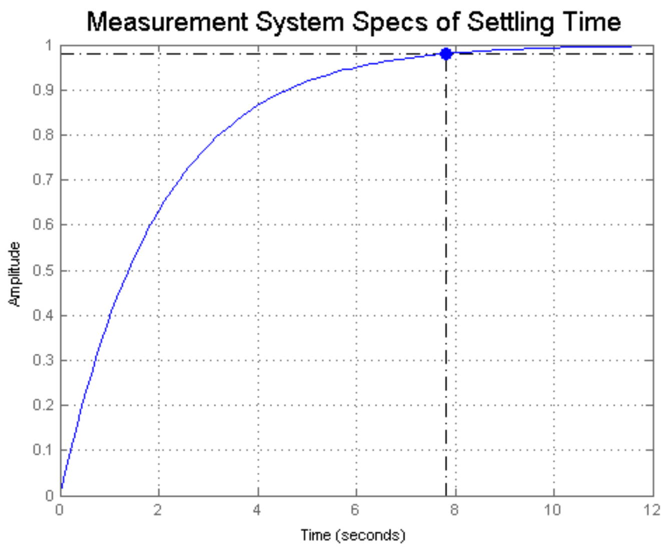
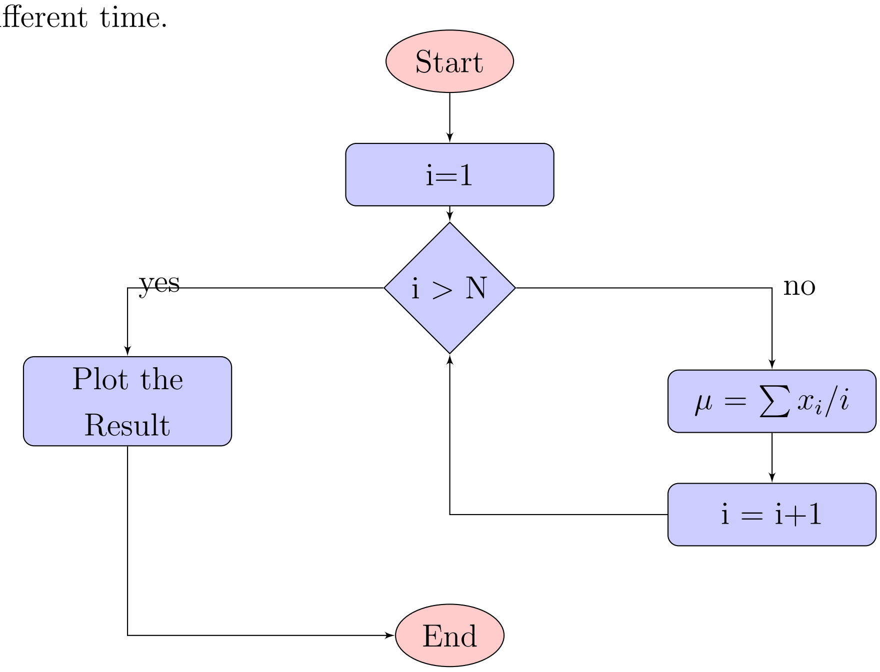


Figure 3: Step Response of Measuring System

In this case, the settling time approximate to 8 second.

3.2.2 Eliminate Non-stationary Part

In order to guarantee the data we use are stationary, I test with mean value versus different time.



Where N is Data length and i is number of sample or time of series.

Obviously, we find that when time approaches 80 sec, statistical property could be stationary.
The result is almost 10 times as time constant of step response. To deal with data carefully and conservatively, we take time-series from 200 sec and make sure we could eliminate non-stationary part.

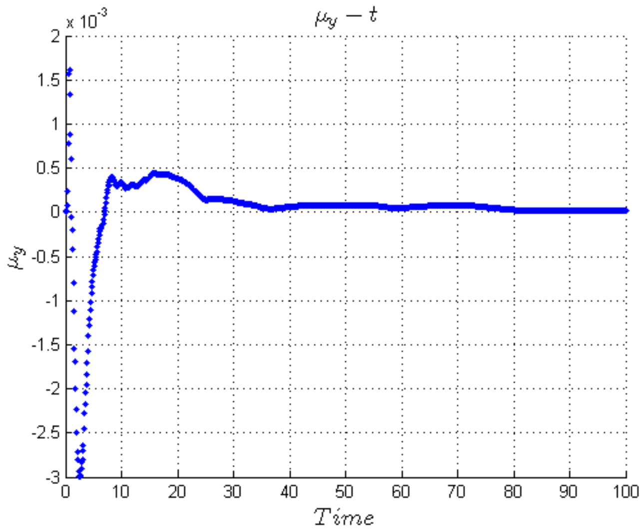


Figure 4: Mean value with Time Series

3.2.3 Compare Ensemble Data with Time-Series Data

I process the data with properties below

Data Type	Ensemble Average	Time Averge
Time span	200 $\sim$ 600 sec	200 $\sim$ 1000 sec
Samples	1,000 samples	Specific one
Data length	8000 (for each)	8000

I calculate the difference between ensemble average and time-series average. Define the error function as

$$e(k) = |\mu(t,k) - \mu(t)|$$

where k is sample index.

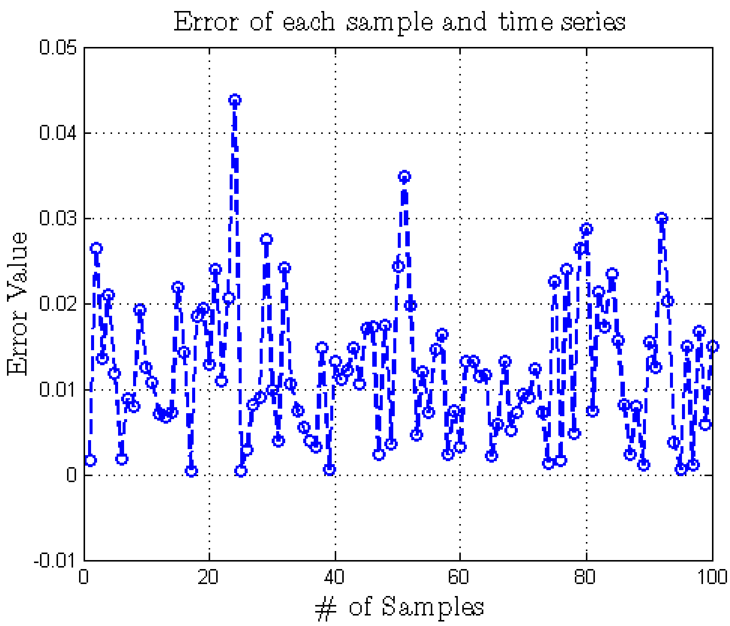


Figure 5: Difference between ensemble average and time-series average

The result shows that the difference is quite small. In engineering, we could find time-series average to substitute ensemble average.

3.2.4 Autocoorelation

We check the autocorrelation of measured data.

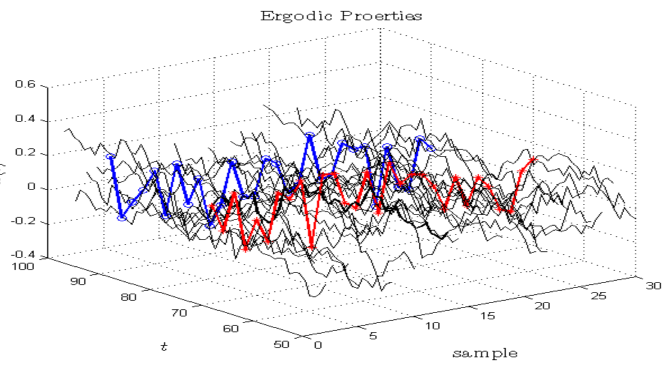


Figure 6: Schematic diagram of autocorrelation

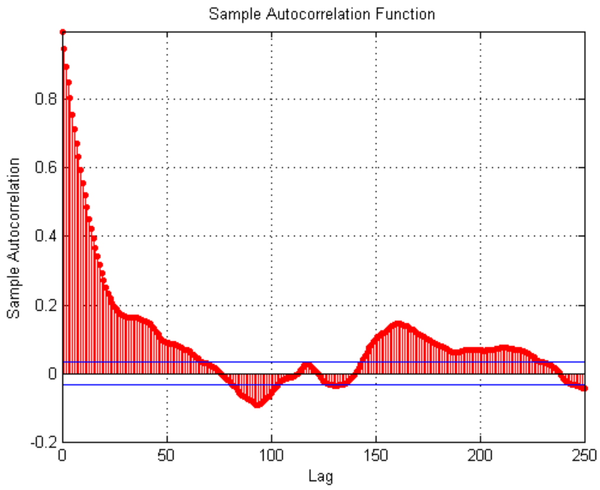


Figure 7: Schematic diagram of autocorrelation

Then the autocorrelation function also shows the envelop decay which indicates

$$\lim_{T\to\infty} \frac{1}{T} \int_{\infty}^{\infty} |R_{xx}(\tau)| d\tau \to 0$$

is one of the property that shows ergodicity.

4 Discussion - Checking if the Data Are Normal

In section 1 we plot histogram to verify the invariant property.

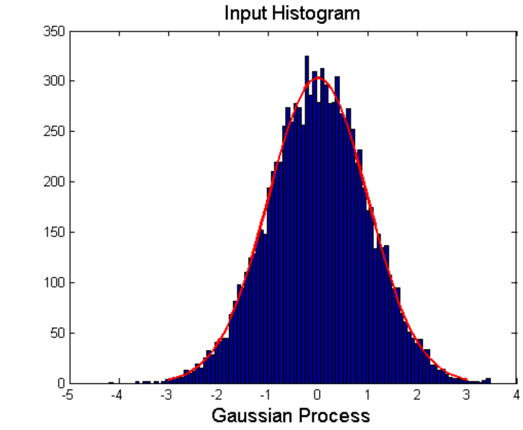


Figure 8: Input Data

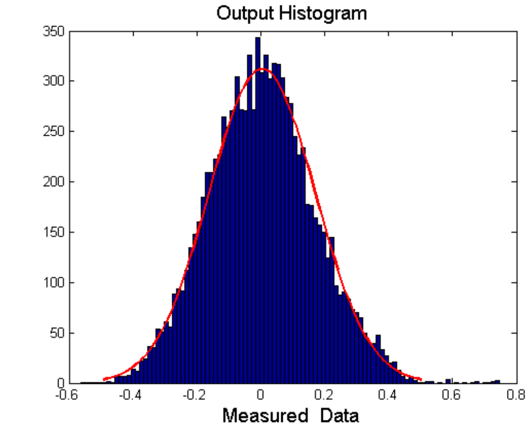


Figure 9: Output Data

But in this way, we check the symmetry by visualization but numerical. If we use the following relation

$$x = \mu + \sigma z$$

Based on the linear relationship which owns slope positive 1, we plot normal scores plot to check the normality.

Figure 10: Input Data

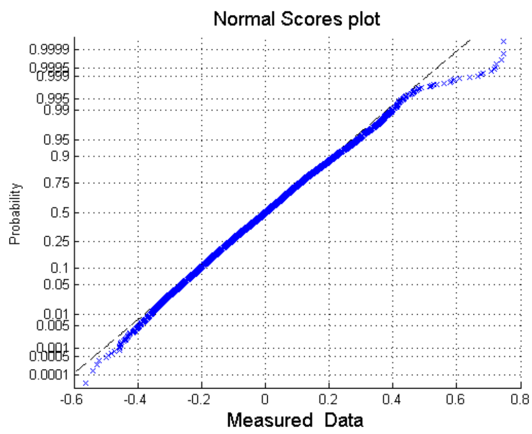


Figure 11: Output Data

So that, we doubly check the Gaussian function possesses the invariant property.

To construct a normal scores plot

- 1. Order the data from smallest to largest
- 2. Obtain the normal scores
- 3. Plot the $i-$th largest observation, versus the $i-$th normal score for all i.

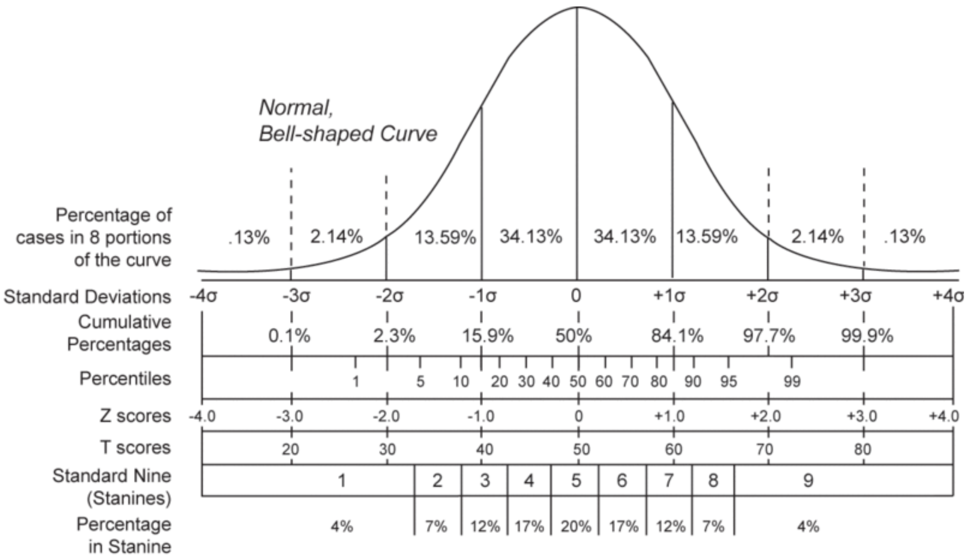


Figure 12: Z- scores

Idealy, we could get the straight line with 45°.

References

[1] Random Data:©: Analysis and Measurement Procedures,Julius S. Bendat, Allan G. Piersol

[2] Wikipedia : Autocorrelation, 2012

[3] Wikipedia : Normal score,Standard score 2012