

Assignment No 3

Tanay Dixit
EE19B123

28-02-2021

Abstract

In this python assignment we study the effects of noise on the fitting process. We analyse the effect of different noise (Gaussian distribution with different σ) applied to the Bessel Function and plot various graphs and analyse the relationship between error in the estimations of functions parameters with changing σ

1 Introduction

We generate a file fitting.dat with 10 columns. The first column is time while the remaining columns are data with a different noise amounts. Noise is assumed to be normally distributed.

The functions used is a linear combination of Bessel Function and a linear function

$$f(t) = A * J_2(t) + B * t + n(t)$$

where A and B are constants and have values of 1.05 and 0.105 respectively. $n(t)$ is the noise function with a given σ and is given by

$$Pr(n(t)|\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-n(t)^2}{2\sigma^2}\right)$$

where sigma are the values varying on a logscale and are obtained by the following function

$$sigma = np.logspace(-1, -3, 9)$$

2 Tasks

2.1 Generation and Loading of Data

The data is generated using the code provided with the assignment. It computes the different values of the sample function using 9 different values of

standard deviation. The data is stored in a .dat file with first column as time values and the rest 9 columns representing the function value with different amount of noise added.

The data is read as a numpy array using `loadtxt()` function

2.2 Plotting of data

We first plot the different noisy functions with time t on x axis and function value on y axis. The true function value also has been plotted in black as the reference.

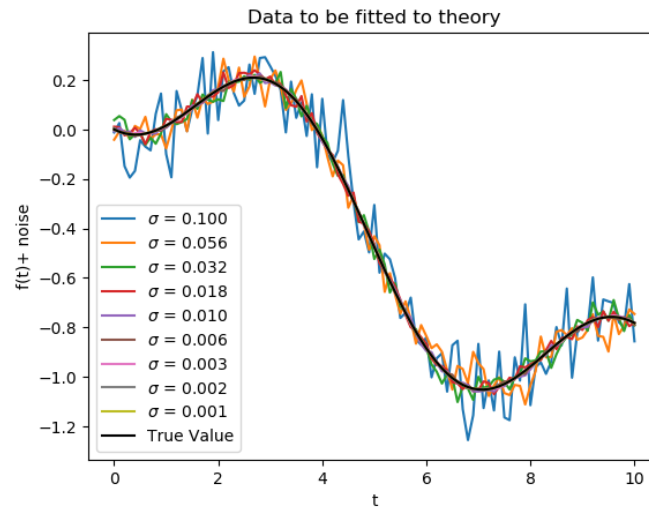


Figure 1: Various Functions

2.3 Error plots

We visualise the error in the measurement using the `errorbar()` function. The graph has been obtained by plotting the first column in the data file which corresponds to $\sigma = 0.1$

The true value has also been plotted for reference. The read bars show the intervals of uncertainty where the noisy data points could lie

Here we plot the error bar for one data column:

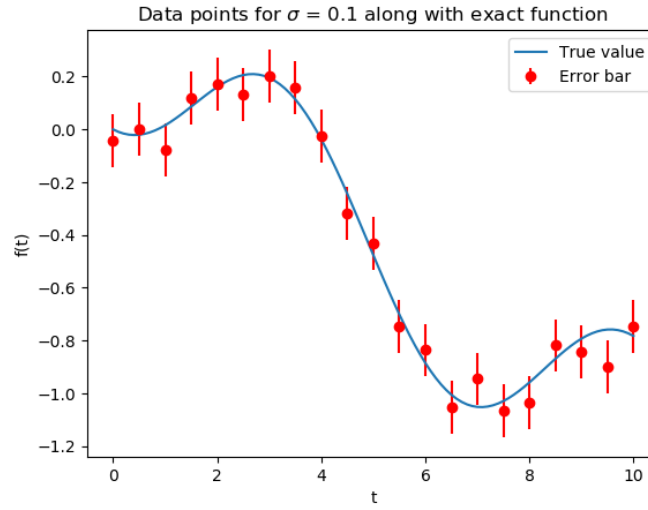


Figure 2: Error bars for $\sigma = 0.1$ along with exact function

2.4 Matrix Generation

We transform this problem into a matrix multiplication task as it gets simpler/faster for computing

$$g(t,A,B) = \begin{pmatrix} J_2(t_1) & t_1 \\ J_2(t_2) & t_1 \\ \dots & \dots \\ J_2(t_n) & t_n \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = M \cdot p$$

where M matrix are the values in the data file and p matrix is the functions parameters

NOTE: a simple test is performed to check if this matrix multiplication is consistent with the standard approach.

2.5 Contour plots of error function

For the given function

$$g(t, A, B) = AJ_2(t) + Bt$$

we vary the function parameters (A and B) on a linear scale and plot the error matrix for all combinations of A and B.

The general error matrix is given by

$$\epsilon_{ij} = \frac{1}{101} \sum_{n=0}^{101} (f_k - g(t_k, A_i, B_j))^2$$

We then plot the contour plot of ϵ_{ij}

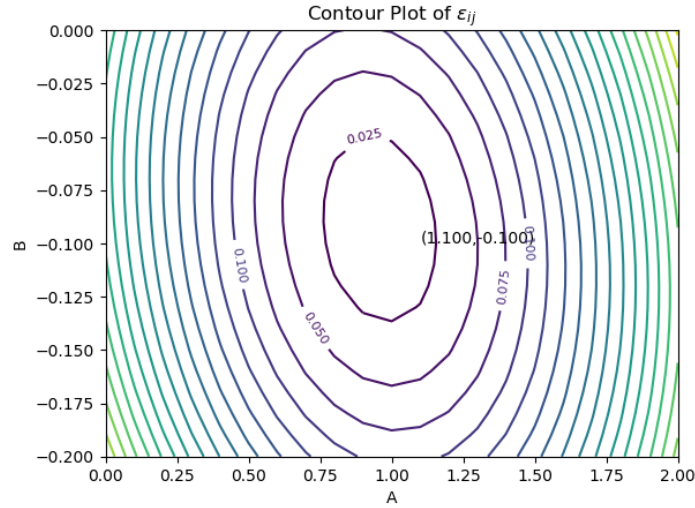


Figure 3: contour plot for ϵ_{ij}

From the above plot we can clearly see that there exist a single minimum.

2.6 Estimations

For estimating the values of A and B we solve for

$$\operatorname{argmin}((M \cdot AB - b)^2)$$

we make use of the `scipy.linalg.lstsq()` function

```
def ls_estimate(M,p):  
    return linalg.lstsq(M,p)[0]
```

2.7 Error plots for different scales

2.7.1 Linear scale

We plot the graph of error with changing values of σ on a linear scale

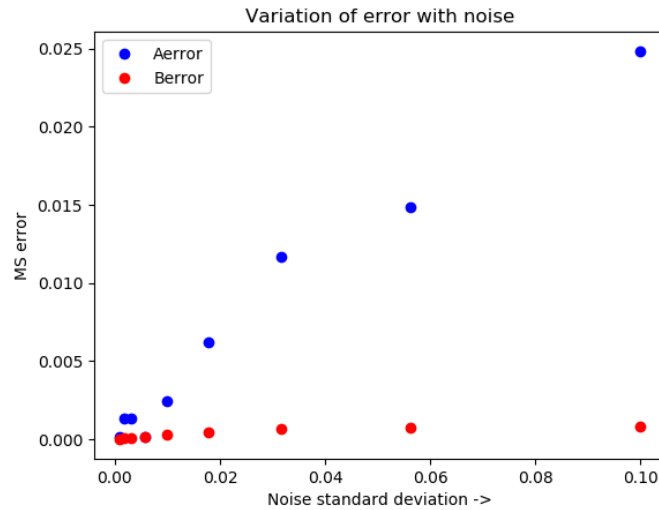


Figure 4: A and B error in linear scale

The errors in the estimates of A and B are non-linear with respect to σ on the linear scale

2.7.2 Log scale

Now we plot the graph in log scale

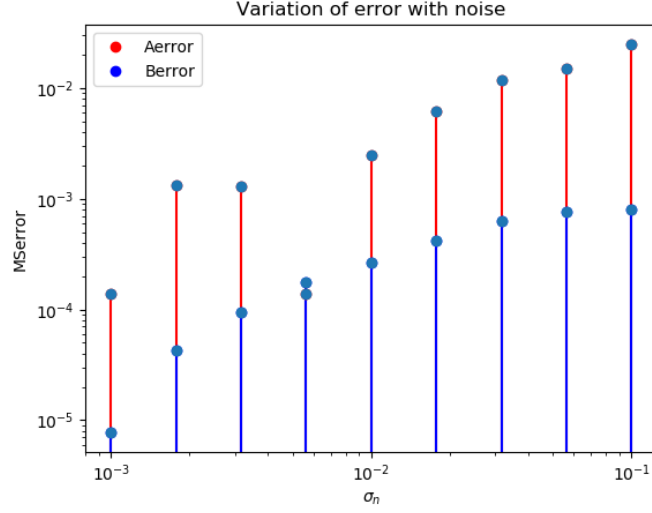


Figure 5: A and B error in log scale

Here we can see the relationship is almost linear as expected, this is true because sigma varied on a log scale and hence the error was also expected to vary so as its a gaussian function.

3 Conclusion

For the given noisy function the best possible estimates for A and B were obtained by minimizing the mean squared error with the true value. It is observed that error in estimation of the model parameters changes approximately linearly with respect to σ in the log scale