Project 1

Shaotian Wu, Chenting Zhang

1. INTRODUCTION

In this report, we will discuss some properties of the Gaussian Distribution and system models with Gaussian Noise. We partitioned this project into two parts. In part A, we will analyze the impact of the length of sequences on the approximation of original distribution and the impact of the correlation coefficient on the shape of the joint pdf of two-dimensional Gaussian variables. In part B, …..

1. PROBLEM FORMULATION AND SOLUTION
   1. Part 1

Task 1: We denote n as the length of sequence x(n), as the value of x at the index of i, as the mean of the sequence x(n), as the variance of the sequence x(n).

For discrete random variables, we could calculate the mean of the variance of each sequence by using the formula as follows:

, .

Then we could calculate the estimated mean and variance of three sequences respectively, the results are shown in the table as follows:

|  |  |  |
| --- | --- | --- |
| } | Mean | Variance |
|  | 1.383 | 5.638 |
|  | 0.613 | 2.171 |
|  | 0.441 | 1.960 |

Table 1: The Mean and Variance. Figure 1: Empirical CDFs of three sequences

The images of each empirical distribution sequence could be drawn as follows:

We could conclude from Figure 1 that the estimation of distribution parameters becomes more accurate and the corresponding empirical distribution is much closer to the real Gaussian distribution when the sample size increases.

Task 2: We denote , as the mean and variance of the Gaussian distribution X, , as the mean and variance of the Gaussian distribution Y. is the correlation coefficient between X and Y. For joint Gaussian distribution , the two-dimensional probability density function of a vector [x, y] could be written as follows:

The sign of the correlation coefficient indicates the direction of the linear relationship between x and y. When is near 1 or −1, the linear relationship is strong; when it is near 0, the linear relationship is weak.

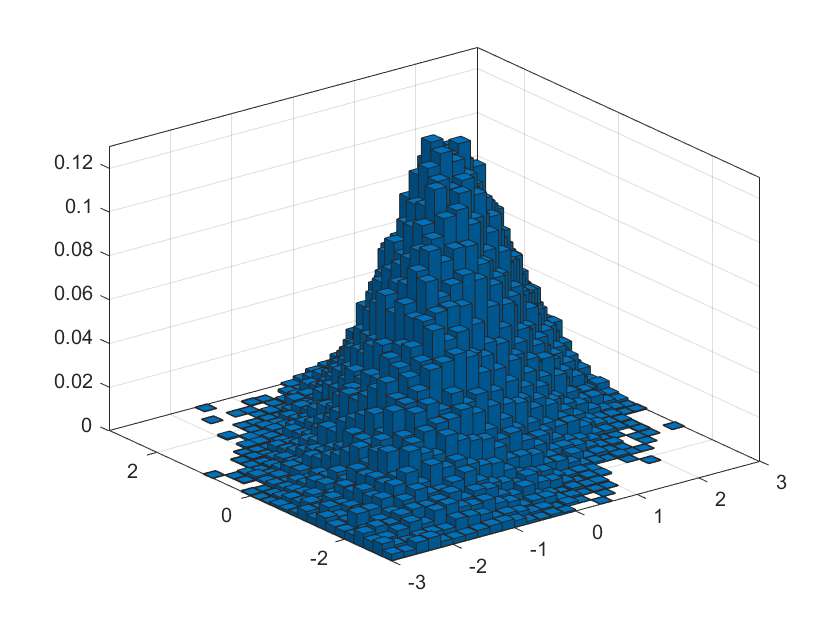
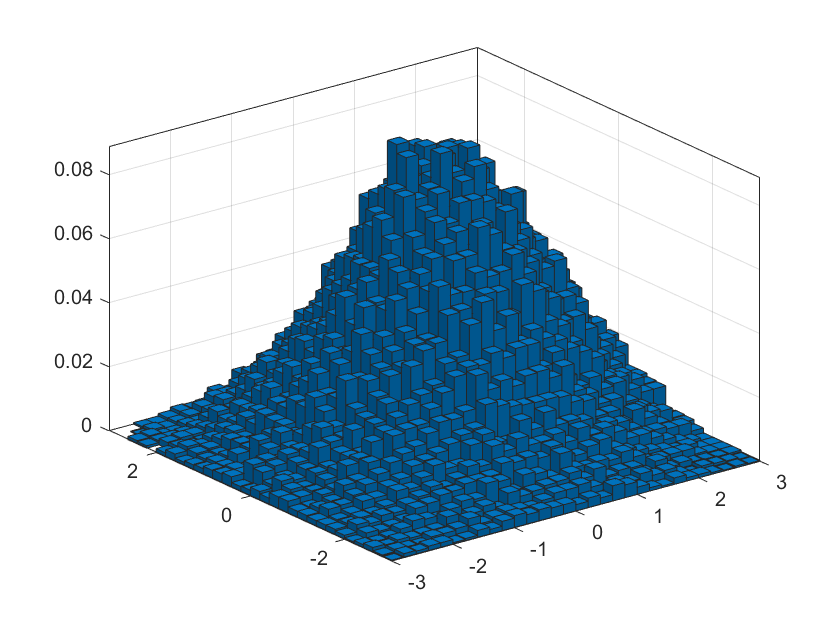


Figure 2: Empirical pdfs of two sets of two-dimensional matrices

We could observe from Figure 2 that linear relationship between X and Y in sequence2 is strong while in sequence1 is weak, so the correlation coefficient of X and Y is 0.25 and 0.75 respectively. Apart from that, if the correlation coefficient is less than 0, a negative correlation occurs and both variables move in the opposite direction. In other words, it means that if one variable increases, the other variable would decrease, vice versa.

Task 3: The pdf of one-dimensional Gaussian variable is written as follows:

Then by the definition of joint distribution in formula (1), the closed-form of the conditional pdf with random variable could be written as follows:

A linear transformation of a Gaussian distribution is still a Gaussian, and also follow the Gaussian distribution. Assume that the variances of X and Y are equal to . The mean and variance of Z=X+Y is derived as follows:

Let , , then

.

.

So, ,

Then plug in the mean and variance into formula(2) respectively, we could obtain the pdfs of these two random variables.

.

1. Part 2

Task 4:

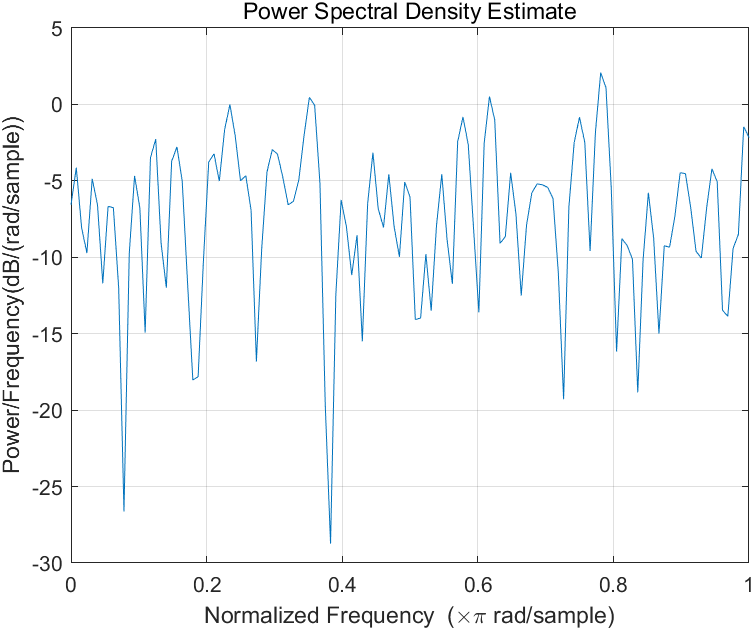
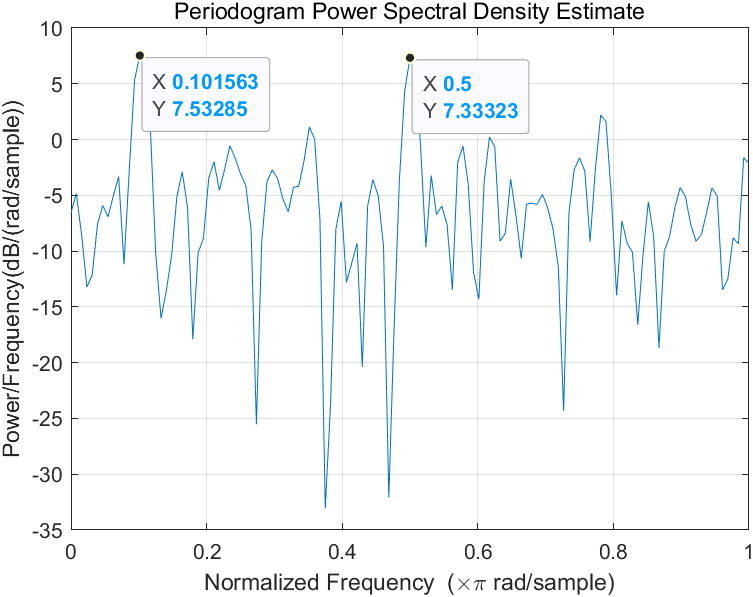
 

Figure 3: The periodograms of signal y0 and y1

The periodograms of the two signals are shown above. Periodogram Power Spectral Density Estimate Power/Frequency Normalized Frequency

In order to match the plots between H0 and H1, we must take a look at the difference between the two plots. Obviously, in the plot of y1, there are two peaks that do not appear in the plot of y0. The normalized frequency of the two peaks is shown in the plot. Hence, H0 should be matched with y0, which is a white noise sequence and H1 should be matched with y1, which is a white noise sequence with two sinusoids attached to it.

Through the statistics of the two peaks in plot of y1 we can see that the peak’s normalized frequency, 0.101 and 0.5, are the double of given v0 and v1. This is because that in the periodogram plotting the Nyquist normalized frequency is used for plotting. If we are to have the actual cut off frequency, which is the actual frequencies of the two sinusoids, we need to half the frequency that in the peak. So the actual frequency of the two sinusoids are

The two results are nearly equal to the given frequencies. Thus, the accuracy of the two results are ensured.

Task 5: The periodogram of the colour-noised signal is shown below.

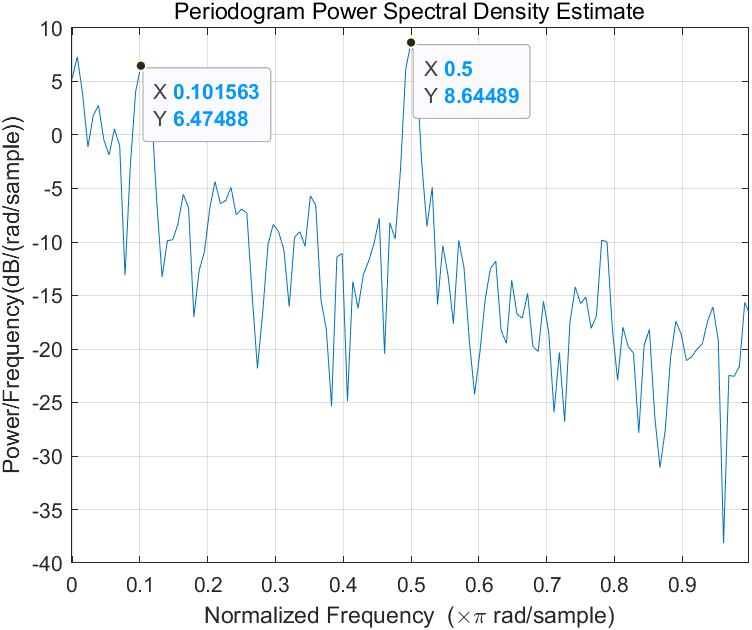


Figure 4: The periodogram of signal y, which is colour-noised

When the noise is coloured, it means that the estimation would not be unbiased and we can see obviously from the plot that the estimation is not unbiased. The two peaks for the two sinusoid still remains but as the frequency goes up the power would obviously get biased.

Task 6, Task 7: According to the given model, , where z(n) is a white noise. We can know that this process is autoregressive(AR) process. This lead to

Future, for ,

So we can get the power spectra

.

Because and , .

For , , where is the mode of the system’s frequency. Given ,. As a result,

.

In order to derive , we need to do the inverse Discrete Fourier Transform to . Thus, . The plots of , and are shown below.

Figure 5,6,7: The plots of , and

1. CONCLUSIONS
2. The estimation of distribution parameters becomes more accurate and the corresponding empirical distribution is much closer to the real Gaussian distribution when the sample size increases.
3. The sign of the correlation coefficient indicates the direction of the linear relationship between x and y. When is near 1 or −1, the linear relationship is strong; when it is near 0, the linear relationship is weak.
4. We could draw from Image 3 and Image 4 that H0 should be matched with y0 and H1 should be matched with y1, and it is a effective way to estimate the frequency of a sinusoid in periodogram if the sinusoid is corrupted by a white noise. A coloured noise, however, will do harm to the recognition of the sinusoid because of biased estimation.
5. When calculating the power spectra of a discrete sequence, we need to calculate its ACF first, and use Fourier Transform to get the power spectra of the sequence. The power spectra of the signal that enters a linear process is the mode of the process’s square in frequency field multiplied with the original function’s ACF.

REFERENCES

1. P. Handel, R. Ottoson, H. Hjalmarsson, Signal Theory, KTH, 2012
2. MATLAB manual for periodogram function, https://ww2.mathworks.cn/help/signal/ref/periodogram.html#bufqp9w