Project 1

Shaotian Wu, Chenting Zhang

I. 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, we will analyze the impact of the noise attached to sinusoid signals based on periodogram and derive and plot the power spectrum and ACF of AR(1) signals.

II. PROBLEM FORMULATION AND SOLUTION

A. Part 1

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

$$m_x = E(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \ , \ var_x = \frac{1}{n}\sum_{i=1}^{n} \ (x_i - m_x)^2.$$

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

$\{x_i(n)\}_{i=1}^{3}$	Mean	Variance
$x_1(n)$	1.383	5.638
$x_2(n)$	0.613	2.171
$x_3(n)$	0.441	1.960

Table 1: The Mean and Variance.

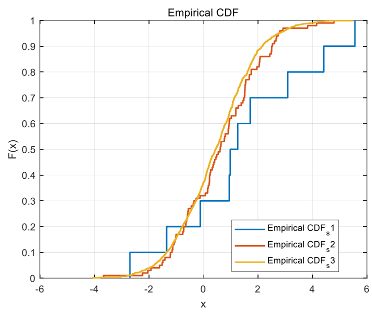

Figure 1: Empirical CDFs of three sequences

The images of each empirical distribution sequence could be drawn as above.

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 u_1, σ_1 as the mean and variance of the Gaussian distribution X, u_2, σ_2 as the mean and variance of the Gaussian distribution Y. The edge distribution of a two-dimensional normal distribution is a one-dimensional normal distribution. The parameter r represents the correlation coefficient between X and Y. ρ is the correlation coefficient between X and Y. For joint Gaussian distribution $f_{XY}(x, y)$, the two-dimensional probability density function of a vector [x, y] could be written as follows:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-u_1}{\sigma_1}\right)^2 - 2r + \left(\frac{y-u_2}{\sigma_2}\right)^2\right]\right\}. \quad (1)$$

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

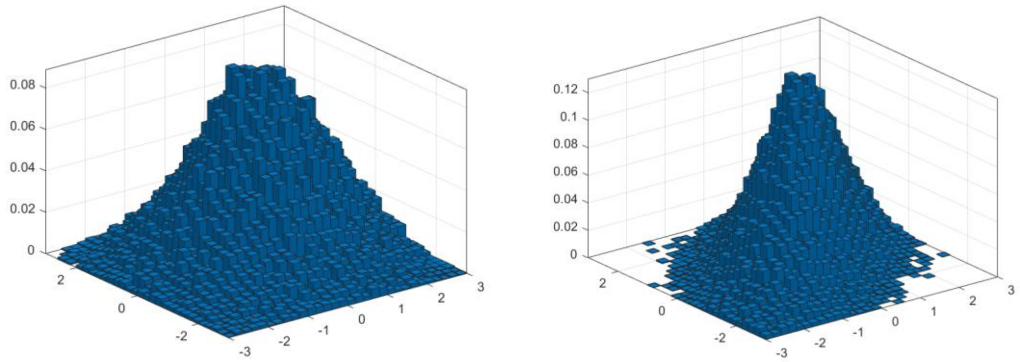

Figure 2: Empirical pdfs of two sets of two-dimensional matrices

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 ρ 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:

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-u_2)^2}{2\sigma^2}\right\}. \quad (2)$$

Then by the definition of joint distribution in formula (1), the closed-form of the conditional pdf $f_Z(z)$ with random variable $Z = X|Y$ could be written as follows:

$$f_z(\mathrm{z}) = f_{X|Y=y}(\mathrm{x}) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \exp\left\{-\frac{\left[(x-u_X)-\rho\left(y-u_y\right)\right]^2}{2\sigma^2(1-\rho^2)}\right\}.$$

A linear transformation of a Gaussian distribution is still a Gaussian, so $Z = X + Y$ and $Z = X - Y$ also follow the Gaussian distribution. Assume that the variances of X and Y are equal to σ^2. The mean and variance of Z=X+Y is derived as follows:

$$E(X \pm Y) = E(X) \pm E(Y) = u_1 \pm u_2,$$

$$\begin{aligned} D(X \pm Y) &= E[(X \pm Y) - E(X \pm Y)]^2 = E[(X - EX) \pm (Y - EY)]^2 \\ &= E[(X-EX)^2 \pm 2(X-EX)(Y-EY) + E(Y-EY)^2] \\ &= E(X-EX)^2 \pm 2E(X-EX)(Y-EY) + E(Y-EY)^2 \\ &= DX + DY \pm 2Cov(X,Y) = 2\sigma^2 \pm 2Cov(X,Y), \end{aligned}$$

$$Cov(X,Y) = E(X-u_1)(Y-u_2) = \iint_{-\infty}^{+\infty} (x-u_1)(y-u_2)f(x,y)dxdy = \rho\sigma^2.$$

$$D(X \pm Y) = 2\sigma^2 \pm 2Cov(X,Y) = 2\sigma^2 \pm 2\rho\sigma^2.$$

So, $X+Y \sim N(u_1+u_2, 2\sigma^2+2\rho\sigma^2)$, $X-Y \sim N(u_1-u_2, 2\sigma^2-2\rho\sigma^2)$.

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

$$f(Z|z=x+y) = \frac{1}{\sqrt{2\pi(2\sigma^2+2\rho\sigma^2)}} \exp\left\{-\frac{[z-(u_1+u_2)]^2}{2(2\sigma^2+2\rho\sigma^2)}\right\},$$

$$f(Z|z=x-y) = \frac{1}{\sqrt{2\pi(2\sigma^2-2\rho\sigma^2)}} \exp\{-\frac{[z-(u_1-u_2)]^2}{2(2\sigma^2-2\rho\sigma^2)}\}.$$

B. Part 2

Task 4: There are two different signals, H0 and H1, a white noise signal and a white noise signal added with two sinusoid signals. The two signals' output sequences, y0 and y1, are given. By plotting the periodogram we need to figure out which given output sequence applies to which signal. The blue line stands for y0 and the red line stands for y1.

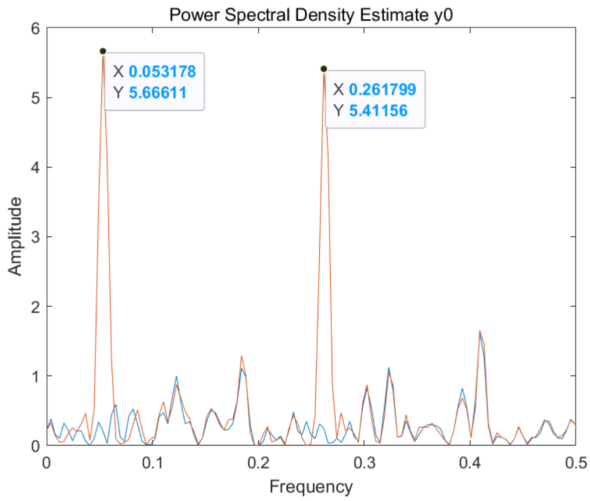

Figure 3: The periodograms of signal y0 and y1

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 frequency are 0.053 and 0.262. The two results are nearly equal to the given frequencies. Thus, the accuracy of the two results is ensured.

Task 5: As for the given sinusoid signals, now we attach a coloured noise to it. We need to analyze the difference through periodogram between the signal with white noise and coloured noise. The periodogram of the colour-noised signal is shown in Figure 4.

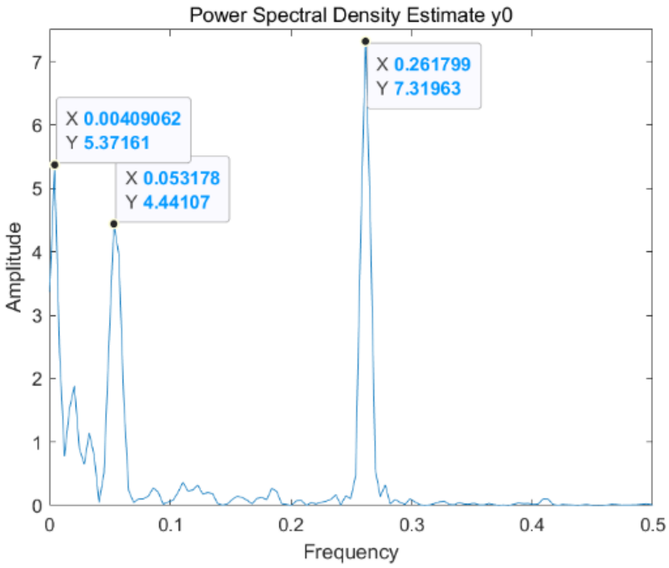

Figure 4: The periodogram of signal y, which is colour-noised

When the noise is coloured, the two peaks of the sinusoid signal has not changed, but there is an extra peak in the figure. The extra peak will bring some difficulties when we are going to recognize the sinusoid signals. Thus, the ability to estimate signal frequencies are worse.

Task 6, Task 7: According to the given model, $X_1(n) = \alpha X_1(n-1) + z(n)$, where z(n) is a white noise. We can know that this process is autoregressive(AR) process. This lead to

$$\gamma_{X_1}(0) = E\left[\left(\alpha X(n-1) + Z(n)\right)^2\right] = \alpha^2 \gamma_X(0) + {\sigma_Z}^2,$$

$$\gamma_{X_1}(0) = \frac{{\sigma_Z}^2}{1-\alpha^2}.$$

Future, for $k > 0$,

$$\gamma_{X_1}(k) = E\left[\left(\alpha X(n-1) + Z(n)\right)X(n-k)\right] = \alpha^2 \gamma_{X_1}(k-1),$$

$$\gamma_{X_1}(k) = \alpha^{|k|} \frac{{\sigma_Z}^2}{1-\alpha^2}.$$

So we can get the power spectra

$$R_{X_1}(v) = F\left(r_{X_1}(\tau)\right) = \frac{\sigma_Z^2}{1-\alpha^2} \cdot \frac{1-\alpha^2}{1+\alpha^2-2\alpha\cos(2\pi v)} = \frac{\sigma_Z^2}{1+\alpha^2-2\alpha\cos(2\pi v)}.$$

Because $\sigma_Z^2 = 1$ and $\alpha = 0.25$, $R_{X_1}(v) = \frac{1}{1.0625-0.5\cos 2\pi v}$.

For $x_2(n)$, $R_{X_2}(v) = |H(v)|^2 R_{X_1}(v)$, where $H(v)$ is the mode of the system's frequency. Given $h(n) = \beta^n u(n)$, $H(v) = \frac{1}{1-\beta e^{-j2\pi v}}$. As a result,

$R_{X_2}(v) = \left|\frac{1}{1-0.25e^{-j2\pi v}}\right|^2 \frac{1}{1.0625-0.5\cos 2\pi v} = \frac{1}{(1-0.25\cos(2\pi v))^2+(0.25\sin(2\pi v))^2} \times$
$\frac{1}{1.0625-0.5\cos 2\pi v} = (\frac{1}{1.0625-0.5\cos(2\pi v)})^2.$

In order to derive $r_{X_2}(k)$, we need to do the inverse Discrete Fourier Transform to $R_{X_2}(v)$. Thus, $r_{X_2}(k) = F^{-1}\left(R_{X_2}(v)\right) = F^{-1}\left((\frac{1}{1.0625-0.5\cos(2\pi v)})^2\right) = \alpha^{|k|}\frac{{\sigma_Z}^2}{1-\alpha^2} * \alpha^{|k|}\frac{{\sigma_Z}^2}{1-\alpha^2} = \frac{256}{225} \times 0.25^{|k|} * 0.25^{|k|} = \frac{256}{225}\sum_{-\infty}^{\infty} 0.25^{|m|}0.25^{|k-m|} = \frac{256}{225}\left(\sum_{-\infty}^{0} 0.25^{|m|}0.25^{|k-m|} + \sum_{1}^{k} 0.25^{|m|}0.25^{|k-m|} + \sum_{k+1}^{\infty} 0.25^{|m|}0.25^{|k-m|}\right) = \frac{256}{225}(\frac{17}{15} + k)(\frac{1}{4})^k$. The plots of $R_{X_1}(v)$, $R_{X_2}(v)$ and $r_{X_2}(k)$ are shown in Figure 5,6,7.

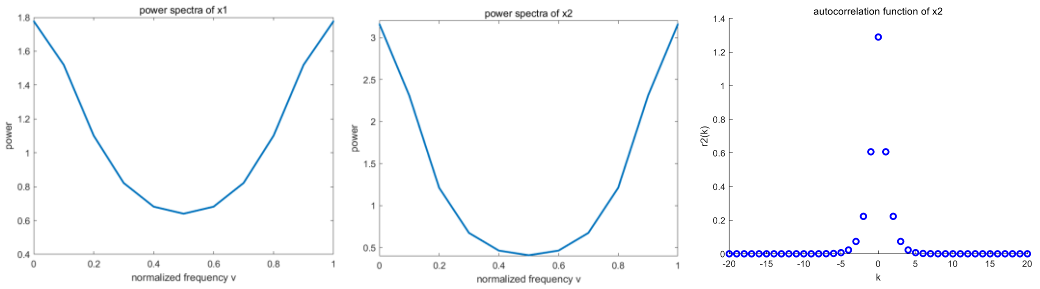

Figure 5,6,7: The plots of $R_{X_1}(v)$, $R_{X_2}(v)$ and $r_{X_2}(k)$

III. CONCLUSIONS

(1) 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.
(2) The sign of the correlation coefficient indicates the direction of the linear relationship between x and y. When ρ is near 1 or −1, the linear relationship is strong; when it is near 0, the linear relationship is weak.
(3) From periodograms sinusoid signals can be extracted when corrupted by white noise by observing the peak, but coloured noise may have extra peak in the periodogram which will influence the ability to recognize the sinusoid signals.
(4) For an AR process, power spectra can be derived by doing FFT to the ACF of AR process, and the convolution in time zone correspond to the product in frequency zone. When we need to derive the ACF of a process, we do DFT to the power spectra of the process.

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