EQ1220 Signal Theory - Project Assignment 1

Yuqi Zheng, Ernan Wang

I. INTRODUCTION

In this project, we experiment with Gaussian distribution parameter estimation, sinusoidal signals with white and colored noise, and AR(k) processes. The project is aimed to enhance our understanding of the concepts and properties related to random processes such as the variance, probability density function, correlation, autocorrelation function, power spectrum, etc. To fulfil our goals, we combine mathematical derivations with the application of MATLAB.

II. PROBLEM FORMULATION AND SOLUTION

1. Task 1

In this task, we need to estimate the mean and variance of three sequences of different lengths N_i, whose elements are i.i.d and follow a Gaussian distribution, $\mathcal{N}(0.5, 2)$.

Similar to a continuous process, the estimators for the mean and variance from a discrete sequence $X = \{x(n)\}$ are given by

$$\hat{m}_X = \frac{1}{N}\sum_{n=0}^{N-1} x(n), \tag{1}$$

$$\hat{\sigma}_X^2 = \frac{1}{N}\sum_{n=0}^{N-1}(x(n) - \hat{m}_X)^2. \tag{2}$$

Hence, we could obtain an estimate of the mean and variance of the sequences $\{x_i(n)\}_{i=1}^{3}$ as follows.

$$\hat{m}_{S_1} = 1.38,\ \hat{\sigma}_{S_1}^2 = 5.64. \tag{3}$$

$$\hat{m}_{S_2} = 0.61,\ \hat{\sigma}_{S_2}^2 = 2.17. \tag{4}$$

$$\hat{m}_{S_3} = 0.44,\ \hat{\sigma}_{S_3}^2 = 1.96. \tag{5}$$

We can see that the accuracy of S_1 is the worst while that of S_2 and S_3 is much better.

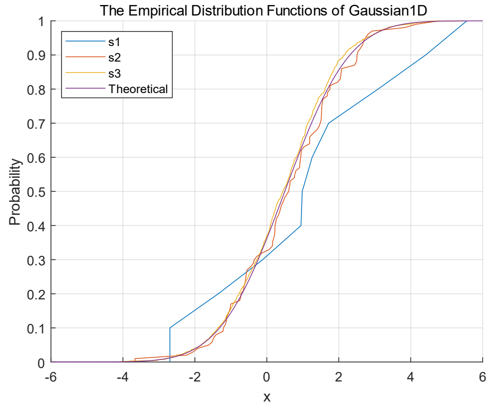

Figure 1:The Empirical cdf of Gaussian1D

Then, we utilize MATLAB to plot the *Empirical Distributions* of the elements as Figure 1.

From the figure, we can see that the curve of S_3 is the closest to the theoretical one, with length $N_3 = 1000$. In contrast, the curve of S_1 is the least similar to the theoretical curve, with length $N_1 = 10$. The curve of S_2 is better than that of S_1 but not as good as the curve of S_3, with a length $N = 100$.

Therefore, we conclude that estimations calculated from a longer sequence are more accurate.

2. Task 2

In this task, we explore the relationship between the correlation coefficient ρ and the pdf of the correlated random variables.

Firstly, the general pdf expression for bivariate Gaussian random variables is given by

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-m_X}{\sigma_X}\right)^2 - \frac{2\rho(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-m_Y}{\sigma_Y}\right)^2\right]}. \tag{6}$$

Especially, in this task, $\sigma_X = \sigma_Y = \sigma$.

In MATLAB, we utilize the function Cov to obtain the covariance matrix directly. Then, we can figure out the correlation coefficient ρ as

$$\rho_{S_1} = 0.2467 \approx 0.25,$$
$$\rho_{S_2} = 0.7518 \approx 0.75. \tag{7}$$

We can utilize the function $ksdensity$ to plot the 3D-plot of the empirical pdfs of the given data in MATLAB as Figure 2.

We can see that the empirical pdf of S_2 with $\rho_{S_2} = 0.75$ is taller and thinner while the opposite is true for S_1 with a smaller correlation coefficient $\rho_{S_1} = 0.25$. This result shows that the greater the absolute value of the correlation coefficient is, the stronger the correlation is, and the more concentrated the elements are.

If we characterize the pdf from $-\rho_i$ instead of ρ_i, the shape of the pdf will not change but the order of the axes will be reversed as Figure 3.

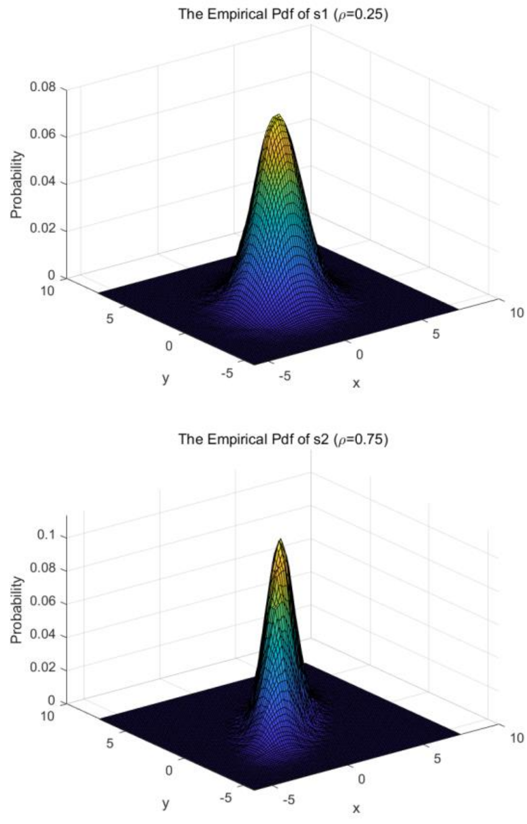

Figure 2: The Empirical Pdfs of S_1 and S_2.

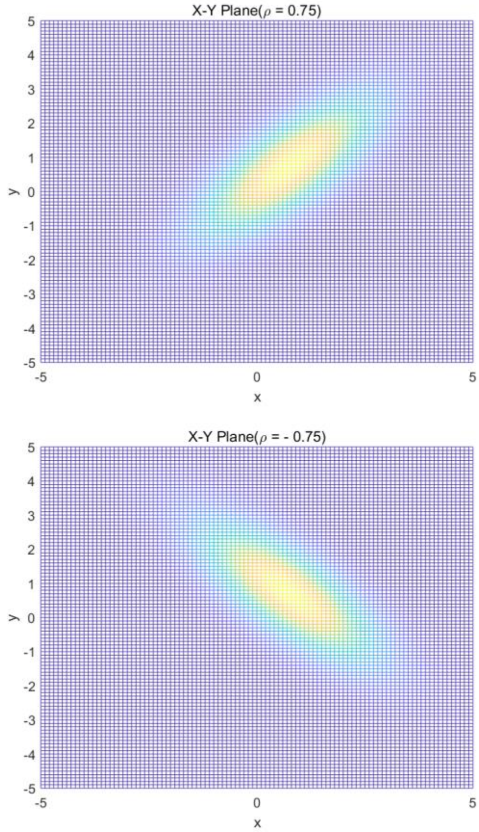

Figure 3: The Empirical Pdfs of S_2 when $\rho_{S_2} = 0.75$ and $\rho_{S_2} = -0.75$.

3. Task 3

In this task, we use the general formula derived in task 2 and the definition of the conditional distribution to derive the conditional pdfs of the random variables $Z = X|Y = y$, $X + Y$, and $X - Y$. To simplify the expressions, we assume that $\sigma_X = \sigma_Y = \sigma$ in this case.

The general expression for the bivariate Gaussian random variables $f_{XY}(x, y)$ is given by (6). According to (2.30) in the textbook[1], the marginal probability density function of Y can be obtained as

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx =$$
$$\int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-m_X}{\sigma}\right)^2 - \frac{2\rho(x-m_X)(y-m_Y)}{\sigma^2} + \left(\frac{y-m_Y}{\sigma}\right)^2\right]} dx =$$
$$\frac{e^{-\frac{(y-m_Y)^2}{2\sigma^2(1-\rho^2)}}}{2\pi\sigma^2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{\left(\frac{x-m_X}{\sigma}\right)^2 - 2\rho\left(\frac{x-m_X}{\sigma}\right)\left(\frac{y-m_Y}{\sigma}\right)}{2(1-\rho^2)}} dx =$$

$$\frac{e^{-\frac{(y-m_Y)^2}{2\sigma^2(1-\rho^2)}}}{2\pi\sigma^2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} e^{-\frac{\left(\frac{x-m_X}{\sigma}-\rho\frac{y-m_Y}{\sigma}\right)^2-\rho^2\left(\frac{y-m_Y}{\sigma}\right)^2}{2(1-\rho^2)}} dx = \frac{e^{-\frac{(y-m_Y)^2}{2\sigma^2}}}{2\pi\sigma^2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} e^{-\frac{\left(\frac{x-m_X}{\sigma}-\rho\frac{y-m_Y}{\sigma}\right)^2}{2(1-\rho^2)}} dx \ . \tag{8}$$

Let $t = \frac{1}{\sqrt{1-\rho^2}}\left(\frac{x-m_X}{\sigma} - \rho\,\frac{y-m_Y}{\sigma}\right)$, therefore,

$$f_Y(y) = \frac{1}{2\pi\sigma}\ e^{-\frac{(y-m_Y)^2}{2\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-m_Y)^2}{2\sigma^2}}, -\infty < y < \infty \ . \tag{9}$$

Hence, we can derive the conditional pdf $f_Z(x)$ of the random variable $Z = X|Y = y$ as follows.

$$f_Z(x) = f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{\dfrac{e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-m_X}{\sigma}\right)^2-\frac{2\rho(x-m_X)(y-m_Y)}{\sigma^2}+\left(\frac{y-m_Y}{\sigma}\right)^2\right]}}{2\pi\sigma^2\sqrt{1-\rho^2}}}{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-m_Y)^2}{2\sigma^2}}} = \frac{1}{\sigma\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-m_X}{\sigma}-\rho\frac{y-m_Y}{\sigma}\right)^2} . \tag{10}$$

Furthermore, $X + Y$ and $X - Y$ are linear combinations of X and Y. According to the properties of the Gaussian distribution[2], $X + Y$ and $X - Y$ also follow Gaussian distributions whose parameters depend on X and Y. The mean and variance are given by

$$m_{X\pm Y} = m_X \pm m_Y, \tag{11}$$

$$\sigma_{X\pm Y}^2 = E\{(X \pm Y - m_{X\pm Y})^2\} = E\{[(X \pm Y) - (m_X \pm m_Y)]^2\} = E\{(X - m_X)^2 + (Y - m_Y)^2 \pm 2(X - m_X)(Y - m_Y)\} = \sigma_X^2 + \sigma_Y^2 \pm 2Cov(X,Y) = \sigma_X^2 + \sigma_Y^2 \pm \rho\sigma_X\sigma_Y = (2 \pm \rho)\sigma^2. \tag{12}$$

The last equation is given by (2.27) in the textbook[1], $Cov(X,Y) = \rho\sigma_X\sigma_Y$. Therefore,

$$f_{X+Y}(z) = \frac{1}{\sqrt{2\pi(2+\rho)\sigma^2}} e^{-\frac{(z-m_X-m_Y)^2}{2(2+\rho)\sigma^2}} , \tag{13}$$

$$f_{X-Y}(z) = \frac{1}{\sqrt{2\pi(2-\rho)\sigma^2}} e^{-\frac{(z-m_X+m_Y)^2}{2(2-\rho)\sigma^2}} . \tag{14}$$

4. Task 4

In this task, we plot the periodogram of the output sequences and match them with the cases H_0 and H_1. We aim to comment on the accuracy of the recovered sinusoidal frequencies.

Figure 4 shows the periodograms of given sequencies y_0 and y_1. From this figure, we can see the psd of y_0 at all frequencies is around 1. Also, there are no dominant frequencies in the periodogram of y_0. Thus, we can speculate that y_0 is white noise and it matches case H_0. Additionally, we can see peak values of the given sequence y_1 appear at 0.05Hz and 0.25Hz, which indicates y_1 has power at 0.05Hz and 0.25 Hz. Also, except for the psd near these two frequencies, the psd at other frequencies of y_1 is almost equal to the psd of y_0, which implies that y_1 contains y_0 (white noise). Hence, we can match y_1 with case H_1, which is the sum of two sinusoidal signals with frequencies 0.05Hz and 0.25Hz and white noise.

Because we can clearly identify two dominant frequencies from the peak value of y_1 from the above periodogram and these two frequencies matches well with the desired frequencies, we can conclude the recovered sinusoidal frequencies are accurate.

5. Task 5

In this task, we use the periodogram to find the impact of noise correlation on the estimation accuracy of the normalized frequencies v_0 and v_1, respectively.

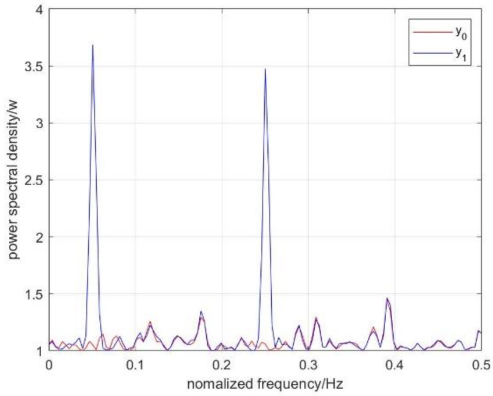

Figure4: Periodograms of sequence y_0 and y_1.

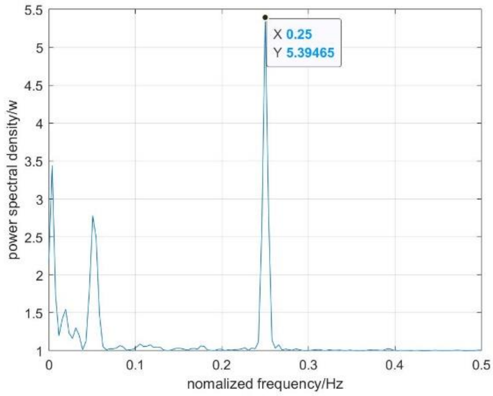

Figure5 : Periodogram of given sequence y.

The periodogram of the given sequence plotted by MATLAB is shown in Figure 5.

From the periodogram of given sequence y, we can see that at 0.25Hz, there is an apparent peak value, while the psd at 0.05Hz has been corrupted by noise. From this, we can see that the correlation noise imposes a greater impact on the estimation of v_0 (0.05Hz) than v_1 (0.25Hz). We can speculate the lower frequencies are easier to be corrupted by correlation noise than higher frequencies.

6. Task 6

In this task, we derive and plot the power spectra $R_{x_1}(v)$ and $R_{x_2}(v)$ of the signals $x_1(n)$ and $x_2(n)$, respectively.

The acf of $x_1(n)$ is listed as below[1]

$$r_{x_1}(k) = \alpha^{|k|} \frac{{\sigma_z}^2}{1-\alpha^2}. \tag{15}$$

The power spectrum can be derived as

$$R_{x_1}(v) = \frac{{\sigma_z}^2}{1-\alpha^2} \sum_{k=-\infty}^{\infty} \alpha^{|k|} e^{-j2\pi v k}$$

$$= \frac{{\sigma_z}^2}{1+\alpha^2-2\alpha\cos(2\pi v)}. \tag{16}$$

Substituting ${\sigma_z}^2$ and α with 1 and 0.25, then we can derive that

$$R_{x_1}(v) = \frac{1}{1.0625-0.5cos(2\pi v)}. \tag{17}$$

From the problem above, we can derive that $x_2(n)$ is the convolution of $h_2(n)$ and $x_1(n)$ as follows,

$$x_2(n) = h_2(n) * x_1(n). \tag{18}$$

The autocorrelation function of $x_2(n)$ can be derived as

$$r_{x_2}(k) = \sum_{l=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} h(l)h(p) r_{x_1}(k + p - l). \tag{19}$$

The power spectrum of $x_2(n)$ can be derived as[1]

$$R_{x_2}(v) = R_{x_1}(v)|H(v)|^2. \tag{20}$$

Since $h_2(n) = \beta^n u(n)$, the discrete time Fourier transform of $h_2(n)$ is listed as below,

$$H(v) = \frac{1}{1-\beta e^{-j2\pi v}}. \tag{21}$$

So $R_{x2}(v)$ can be derived as

$$R_{x_2}(v) = R_{x_1}(v) \left|\frac{1}{1-\beta e^{-j2\pi v}}\right|^2. \tag{22}$$

Substituting $R_{x_1}(v)$ with $\frac{1}{1.0625-0.5\cos(2\pi v)}$ and β with 0.25, then we can derive the equation below,

$$R_{x_2}(v) = \frac{1}{1.0625-0.5\cos(2\pi v)} \left|\frac{1}{1-0.25e^{-j2\pi v}}\right|^2. \tag{23}$$

The plotted power spectra are shown in Figure 6.

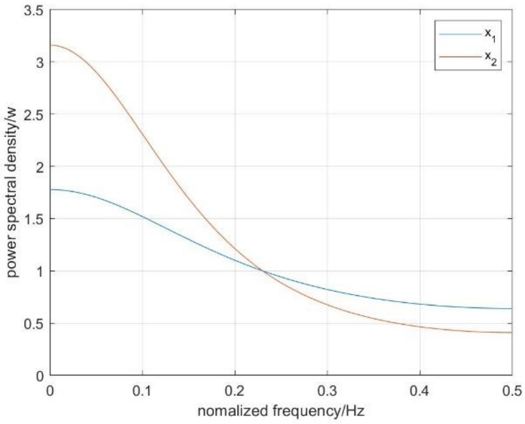

Figure 6: Power spectra $R_{x_1}(v)$ and $R_{x_2}(v)$ of $x_1(n)$ and $x_2(n)$.

7. Task 7

In this task, we derive and plot the acf $r_{x_2}(k)$ of the output process $x_2(n)$. From

$$R_{x_2}(v) = \frac{1}{1.0625-0.5\cos(2\pi v)} \left|\frac{1}{1-0.25e^{-j2\pi v}}\right|^2, \qquad (24)$$

we can derive that

$$r_{x_2}(k) = F_d^{-1}\left[R_{x_2}(v)\right] = \int_{-\frac{1}{2}}^{\frac{1}{2}} R_{x_2}(v) e^{j2\pi vk}\, dv = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{1.0625-0.5\cos(2\pi v)} \left|\frac{1}{1-0.25e^{-j2\pi v}}\right|^2 e^{j2\pi vk} dv. \qquad (25)$$

The acf of $x_2(n)$ plotted by MATLAB is shown in Figure 7.

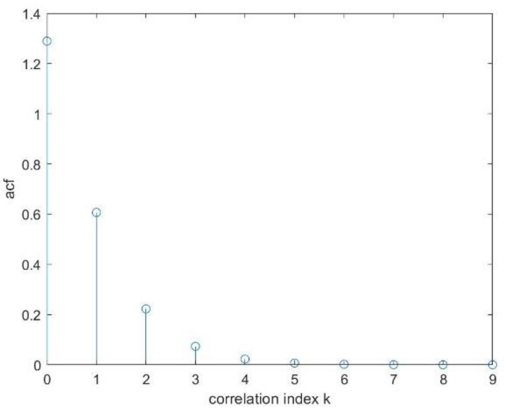

Figure 7: The acf of $x_2(n)$.

III. CONCLUSIONS

In this project, we study Gaussian distribution and system models with Gaussian noise. We find that the length of sampling sequence has an effect on the accuracy of parameter estimation and thus affect the reconstruction of the signal. In addition, through studying the two-dimensional Gaussian distribution, we further understand the significance of the correlation coefficient and the covariance. Besides, we study the effect of Gaussian noise on a sinusoidal signal, which is very important in practical application. Finally, we look into the AR process system model and analyze the relationship between the power spectrum, the autocorrelation function and the transfer function. Actually, they belong to different domains. We are also going to touch the S domain and the Z domain. All of them have important roles to play.

REFERENCES

[1] P. Handel, R. Ottoson, H. Hjalmarsson, Signal Theory, KTH, 2012

[2] Athanasios Papoulis, S. Unnikrishna Pillai, Probability Random Variables and Stochastic Processes