Project 1

Shaotian Wu, Chenting Zhang

I. INTRODUCTION

In this report, we will discuss ~~about~~ some properties of the Gaussian Distribution and system models with Gaussian Noise. ~~The whole passage is partitioned into three parts, introduction, problem formulation and solution, in which we would briefly formulate the problems and provide feasible solutions using mathematical tools. Lastly, we would make a conclusion of this report and highlight the most important parts of the solution.~~

II. PROBLEM FORMULATION AND SOLUTION

A. Part 1

Task 1: Calculate the mean and the variance of the distribution of three sequences, then analyze how the length of sequence impact on the approximation of original distribution.

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

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

Thus, the images of each empirical distribution sequence could be drawn as follows:

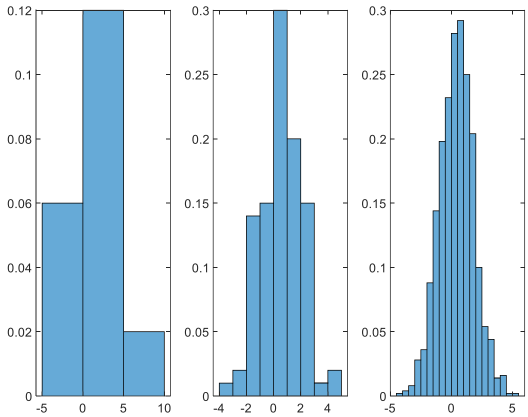

Image 1: Empirical pdfs of three sequenses

We could tell from Image 1 that when N_i increases, the empirical distribution is much closer to the real Gaussian distribution.

Task 2: Write down the general expression for the joint Gaussian distribution $f_{XY}(x, y)$, then plot three-dimensional empirical pdfs of the two matrices. Lastly, analyze the relationship between the shape of pdf and the correlation coefficient.

~~Solution:~~ For joint Gaussian distribution $f_{XY}(x, y)$, the two-dimensional probability density function of a vector [x, y] is written as follows:

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

$\rho = \dfrac{cov(X,Y)}{\sigma_x\sigma_y}$, Where ρ is the correlation between X and Y and where $\sigma_1 > 0$ and $\sigma_2 > 0$

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.

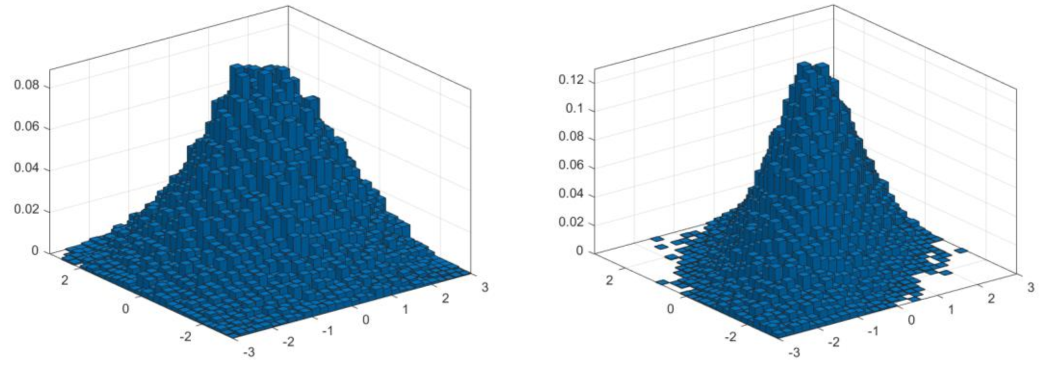

Image 2: Empirical pdfs of two sets of two-dimensional matrices

We could draw from Image 2 that linear relationship between X and Y in sequence2 is strong while in sequence1 is weak, so its correlation coefficient ρ is 0.75 and 0.25 respectively. Apart from that, if the correlation coefficient is less than 0, a negative correlation occurs and both variables move in the opposite direction.

Task 3: Show the mathematical derivation and expression of pdfs of conditional distribution $f(Z = x|Y = y)$, X+Y and X-Y at a given correlation coefficient.

~~Solution:~~ 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 due ~~to~~ the definition of conditional distribution in formula (1), we could do as follows:

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

Since Gaussian Process is a linear operator, $Z = X + Y$ and $Z = X - Y$ also follow the Gaussian distribution. We could obtain the mean and variance of these two random variables:

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

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

$$f(Z|z=x+y)=\frac{1}{\sqrt{4\pi\sigma^2}}\exp\{-\frac{[z-(u_1+u_2)]^2}{4\sigma^2}\},\ f(Z|z=x-y)=\frac{1}{\sqrt{4\pi\sigma^2}}\exp\{-\frac{[z-(u_1-u_2)]^2}{4\sigma^2}\}$$

B. Part 2

Task 4: Plot the periodogram of the output sequences and match the plots with the cases H0; H1 defined in (1). Comment on the accuracy of the recovered sinusoidal frequencies.

~~Solution:~~

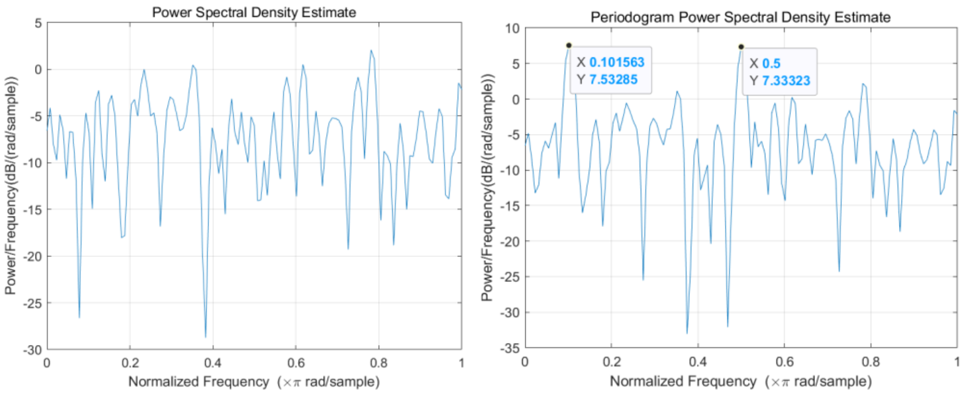

Image 3: The periodograms of signal y0 and y1

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

$$f1=\frac{0.101}{2}=0.051, f2=\frac{0.5}{2}=0.25$$

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

Task 5: Using the periodogram, comment on the impact of noise correlation on the estimation accuracy of the normalized frequencies v0 and v1, respectively.

~~Solution~~: The periodogram of the colour-noised signal is shown below.

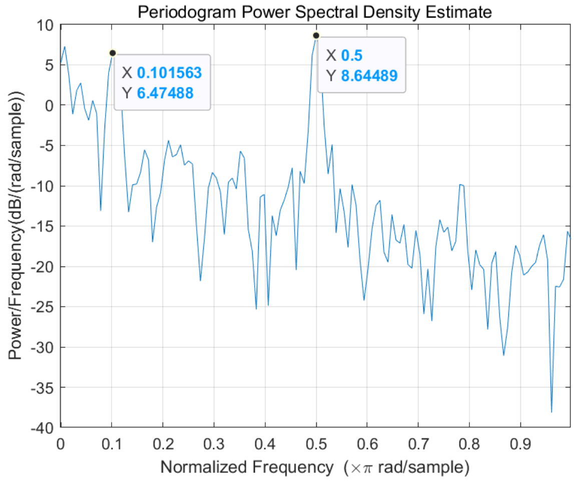

Image 4: The periodogram of signal y, which is colour-noised

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: Derive and plot the power spectra Rx1(v) and Rx2(v) of the signals x1(n) and x2(n), respectively.

~~Solution~~:

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

$$X_1(n) = \alpha X_1(n-1) + z(n) = \alpha\big(\alpha X_1(n-2) + z(n-1)\big) + z(n) + \cdots = \textstyle\sum_{l=0}^{\infty} \alpha^l z(n-l).$$

In order to get the power spectra of X1,we need to calculate the ACF of X1.

$$r_{X_1}(\tau) = E\big(X_1(t)X_1(t+\tau)\big) = E((\alpha X_1(t-1) + z(t))(\alpha X_1(t+\tau-1) + z(t))$$
$$= \alpha^2 r_{X_1}(\tau) + r_z(\tau) + E\big(\alpha X_1(t+\tau-1)z(t)\big) + E(\alpha X_1(t-1)z(t+\tau))$$

$$E\big(\alpha X_1(t+\tau-1)z(t)\big) + E\big(\alpha X_1(t-1)z(t+\tau)\big)$$
$$= E(\alpha \sum_{l=0}^{\infty} \alpha^l z(t+\tau-1-l)z(t)) + E(\alpha \sum_{l=0}^{\infty} \alpha^l z(t-1-l)z(t+\tau))$$
$$= \alpha^l \sigma_z^2$$

As a result, $r_{X_1}(\tau) = \alpha^l \sigma_z^2 + \alpha^2 r_{X_1}(\tau) + r_z(\tau)$. We can get $(1-\alpha^2) r_{X_1}(\tau) = \alpha^l \sigma_z^2 +$ $+ r_z(\tau) = \begin{cases} 2\sigma_z^2, & \tau = 0 \\ \alpha^\tau \sigma_z^2, & \tau \neq 0 \end{cases}$. In order to get the function easy for Fourier Transform, we write the ACF in the linear consistence of the pulse function, so $r_{X_1}(\tau) = \frac{\sigma_z^2}{1-\alpha^2}\delta(\tau) + \frac{\alpha^\tau \sigma_z^2}{1-\alpha^2}$.

So we can get the power spectra $R_{X_1}(v) = F\left(r_{X_1}(\tau)\right) = \frac{\sigma_z^2}{1-\alpha^2}\left(1 + \frac{1-\alpha^2}{1+\alpha^2-2\alpha\cos 2\pi v}\right)$.

Because $\sigma_z^2 = 1$ and $\alpha = 0.25$, $R_{X_1}(v) = \frac{1}{0.9375}\left(1 + \frac{0.9375}{1.0625-0.5\cos 2\pi v}\right)$.

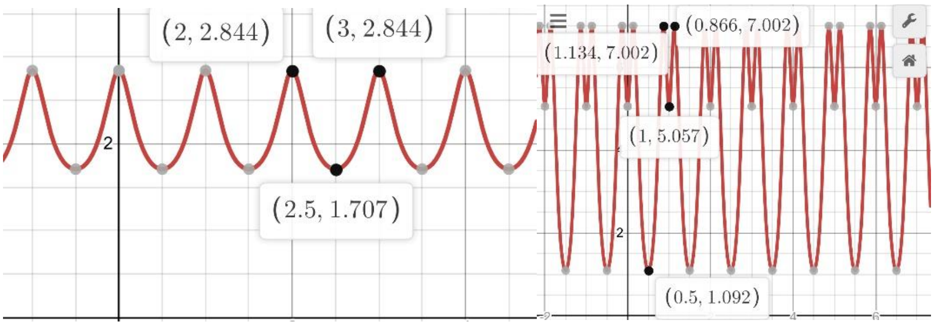

Image 5: The plot of $R_{X_1}(v)$ and $R_{X_2}(v)$

For X2(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, $R_{X_2}(v) = \left|\frac{1}{1-0.25e^{-j2\pi v}}\right|^2 \frac{1}{0.9375}\left(1 + \frac{0.9375}{1.0625-0.5\cos 2\pi v}\right)$.The two functions' plot is shown above.

Task 7: Derive and plot the acf rx2(k) of the output process x2(n).

~~Solution~~: In order to derive the acf $r_{X_2}(k)$, we need to do the inverse Discrete Fourier Transform to $R_{X_2}(v)$. Thus, $r_{X_2}(k) = F^{-1}\left(R_{X_2}(v)\right) = F^{-1}\left(\left|\frac{1}{1-0.25e^{-j2\pi v}}\right|^2 \frac{1}{0.9375}\left(1 + \frac{0.9375}{1.0625-0.5\cos 2\pi v}\right)\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\left|\frac{1}{1-0.25e^{-j2\pi v}}\right|^2 \frac{1}{0.9375}\left(1 + \frac{0.9375}{1.0625-0.5\cos 2\pi v}\right)\right) e^{jvn} dv.$

III. CONCLUSIONS

(1) For the length of Gaussian sequence, When N_i increases, the empirical distribution is much closer to the real Gaussian distribution.

(2) We could draw from Image 2 that linear relationship between X and Y in sequence2 is strong, so its correlation coefficient ρ is 0.75. Likewise, linear relationship between X and Y in sequence1 is weak, so its correlation coefficient ρ is 0.25.

(3) 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.

(4) 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