Signal Theory Project 1

Hanqi Yang, Qian Zhou

I. INTRODUCTION

In this project, we mainly discuss the Gaussian distribution and system models with Gaussian noise. In the first part, we discuss the effect of sequence length on the empirical cdf of a one-dimensional Gaussian distribution and the effect of the correlation coefficient on the pdf of a two-dimensional Gaussian distribution. In the second part, we compare the impact of white noise and coloured noise on the estimation accuracy of the normalized frequencies, and we derive the power spectrum and autocorrelation function of an AR-process.

This report is organized as follows: In section II, some properties of the Gaussian distribution are discussed. Two typical system models with Gaussian noise are displayed in section III. The fourth section contains some conclusions.

II. THE GAUSSIAN DISTRIBUTION

This section contains solutions of 3 tasks.

A. Task 1

To calculate the mean and the variance of the distribution from each sequence $\{x_i(n)\}_{i=1}^{3}$, we use the function *mean()* and *var()* in MATLAB derived from two equations as below.

$$\widehat{m_X} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$\widehat{\sigma_X}^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \widehat{m_X})^2 \tag{2}$$

The results are shown in Table 1.

Table 1: The mean and the variance of $\{x_i(n)\}_{i=1}^{3}$

Estimate value	$x_1(n)$	$x_2(n)$	$x_3(n)$
Mean	1.38	0.61	0.44
Variance	6.27	2.19	1.96

We use the function *ecdf()* to plot the empirical distribution as Figure 1.

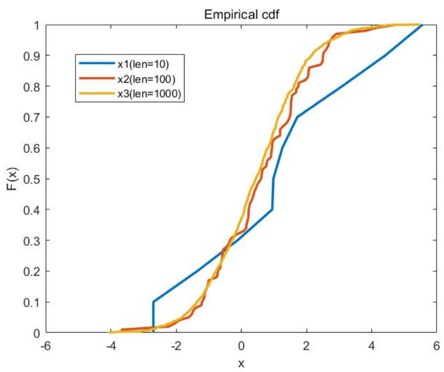


Figure 1: Empirical distribution

As the length of the sequence increases, the more accurate the estimation is and the more the empirical distribution converges to $N(0.5,2)$.

B. Task 2

The general expression for the joint Gaussian distribution from Collection of Formulas in Signal Processing [1] is

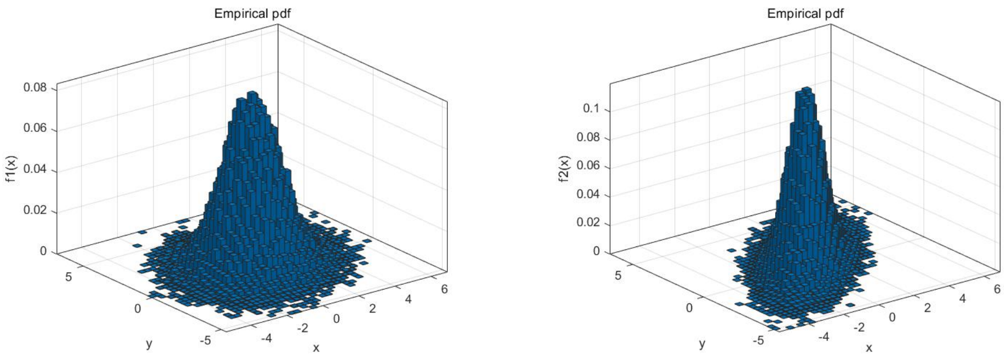
$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{g(x,y)}{2(1-\rho^2)}} \tag{3}$$

where

$$g(x,y)=\frac{(x-m_X)^2}{\sigma_X^2}-\frac{2\rho(x-m_X)(y-m_Y)}{\sigma_X\sigma_Y}+\frac{(y-m_Y)^2}{\sigma_Y^2}$$

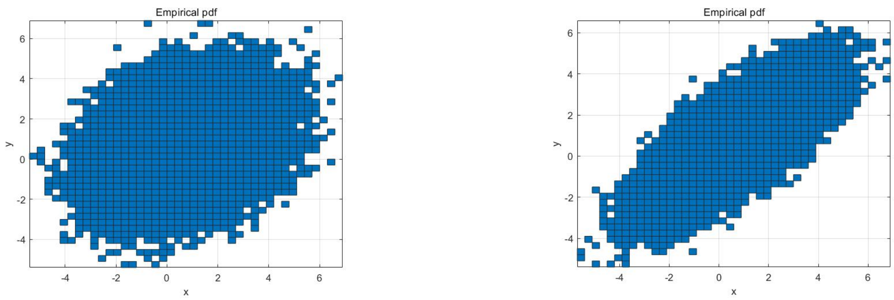
We calculate the mean and the variance of $x_i(n)$ and $y_i(n)$, and use the function *corrcoef()* to derive the estimated correlation coefficient $\widehat{\rho_1}=0.2467$ and $\widehat{\rho_2}=0.7518$. To obtain a 3D-plot of the empirical pdf we use the function *histogram2(X, Y, 'Normalization', 'pdf')* which plots an estimate of the probability density function for X and Y. Hence, the 3D-plots of empirical pdf of $\{x_i(n), y_i(n)\}_{i=1}^{2}$ are shown in Figure 2. $f_{XY}^1(x,y)$ corresponds to $\rho_1=0.25$, and $f_{XY}^2(x,y)$ corresponds to $\rho_2=0.75$.

To better see the effect of the correlation coefficient on the shape of the empirical pdfs, we observe the top view of its 3D-plot, as shown in Figure 3. For ρ_1 and ρ_2, the larger the correlation coefficient, the closer the shape of the pdf is to a line in the XY direction.



(a)$f_{XY}^1(x,y)$ (b)$f_{XY}^2(x,y)$

Figure 2: 3D-plot of the empirical pdfs



(a)$f_{XY}^1(x,y)$ (b)$f_{XY}^2(x,y)$

Figure 3: Top view of 3D-plot of the empirical pdfs

When the correlation coefficient is negative $(-\rho_i)$, X and Y are negatively correlated, i.e. their pdf are mirrored about the y-axis compared to when the correlation coefficient is positive $(+\rho_i)$.

C. Task 3

Both marginal distributions of the 2-D Gaussian distribution are 1-D Gaussian distributions, and neither depends on the correlation coefficient ρ [2]. Therefore, the marginal distribution $f_Y(y)$ of $f_{XY}(x,y)$ is

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{(y-m_Y)^2}{2\sigma_Y^2}} \tag{4}$$

Using the definition of the conditional distribution

$$f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} \tag{5}$$

we can substitute equation (1) and (2) into (3). Since random variables X and Y have equal variance $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, so we can obtain

$$f_{X|Y=y}(x) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{1-\rho^2}} e^{-\frac{g(x,y)}{2(1-\rho^2)}+\frac{(y-m_Y)^2}{2\sigma^2}} \tag{6}$$

A linear combination of two Gaussian distributions still obeys the Gaussian distribution [2]. Hence, the mean m_M and the variance σ_M^2 for M=X+Y is

$$m_M = m_X + m_Y \tag{7}$$

$$\sigma_M^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y = 2\sigma^2(1+\rho) \tag{8}$$

Hence, the pdf for M=X+Y is

$$f_M(m) = \frac{1}{\sqrt{4\pi\sigma^2(1+\rho)}} e^{-\frac{(m-m_X-m_Y)^2}{4\sigma^2(1+\rho)}} \tag{9}$$

Similarly, the the mean m_N and the variance σ_N^2 for N=X-Y is

$$m_N = m_X - m_Y \tag{10}$$

$$\sigma_N^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y = 2\sigma^2(1-\rho) \tag{11}$$

Hence, the pdf for N=X-Y is

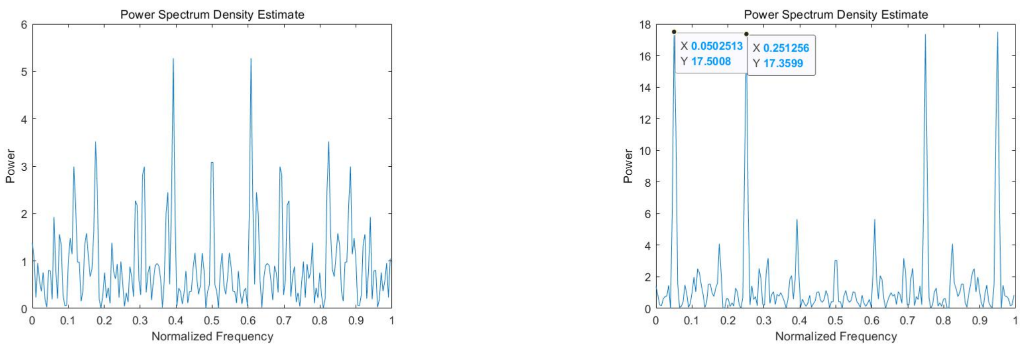
$$f_N(n) = \frac{1}{\sqrt{4\pi\sigma^2(1-\rho)}} e^{-\frac{(n-m_X+m_Y)^2}{4\sigma^2(1-\rho)}} \tag{12}$$

III. SYSTEM MODELS WITH GAUSSIAN NOISE

This section contains solutions of 4 tasks.

A. Task 4

We use the function *xcorr()* to calculate the autocorrelation function of the given data, perform a FFT on the autocorrelation function, then normalise the frequency and finally plot its periodogram in Figure 4. As can be seen from Figure 4(b), there are two dominant frequencies $\hat{\nu_1} = 0.0503$ and $\hat{\nu_2} = 0.2513$, corresponding to the two frequencies $\nu_1 = 0.05$, $\nu_2 = 0.25$ of H1 respectively. Whereas 4(a) does not have a dominant frequencies, it can be considered as white noise, corresponding to H0. Meanwhile, the recovered sinusoidal frequency is very close to the given frequency with deviation 0.6% and 0.52% respectively.



(a)$y_1(n)$ (b)$y_2(n)$

Figure 4: Periodogram of the given sequence in SinusInNoise1.mat

B. Task 5

Using the same method in Task 4, plot the periodogram for the sequence $y(n)$ as Figure 5. It can be seen that $\hat{\nu_0}$ is no longer the dominant frequencies due to the interference of coloured noise. When noise is correlated, the frequency points $\widehat{\nu_0}$ become inaccurate. However, as this coloured noise decays with time, the frequency point $\widehat{\nu_1}$ is still accurate.

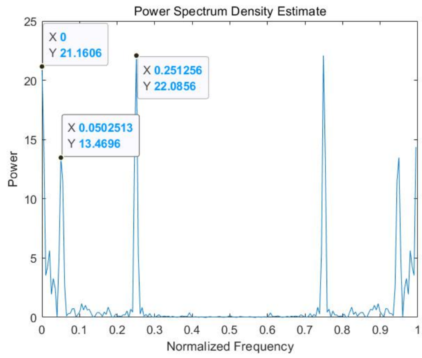


Figure 5: Periodogram of the given sequence $y(n)$ in SinusInNoise2.mat

C. Task 6

Since $x_1(n)$ is an AR process, we use equation (3.36) in Signal Theory [3], substitute $\alpha = 0.25$, $\sigma_z^2 = 1$into (3.36), and derive

$$r_{x_1}(k) = \alpha^{|k|}\frac{\sigma_z^2}{1-\alpha^2} = \frac{4^{2-|k|}}{15} \tag{13}$$

The power spectrum of ARMA-process can be obtained from equation (8.26) in Signal Theory [3]. Then we substitute parameters into (8.26) and get

$$R_{x_1}(\nu) = |H(\nu)|^2\sigma_Z^2 = \left|\frac{1}{1-\alpha e^{-j2\pi\nu}}\right|^2 \sigma_Z^2 = \frac{1}{(1-0.25e^{-j2\pi\nu})^2} \tag{14}$$

We plot $R_{x_1}(\nu)$ in MATLAB in Figure 6, which is the power spectrum of signal $x_1(n)$. According to the question, we can get

$$x_2(n) = h_2(n) * x_1(n) \tag{15}$$

To calculate the power spectrum $R_{x_2}(\nu)$ of signal $x_2(n)$, we use equation (8.14) in Signal Theory [3], and then we can get

$$R_{x_2}(\nu) = R_{x_1}(\nu)H_2(\nu){H_2}^*(\nu) = R_{x_1}(\nu)|H_2(\nu)|^2 \tag{16}$$

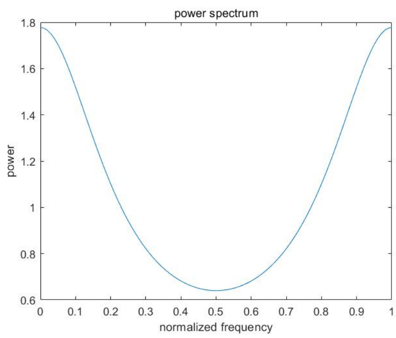
where $H_2(\nu)$ is the Fourier Transform of $h_2(n)$.

$$H_2(\nu) = \frac{1}{1-0.25e^{-j2\pi\nu}} \tag{17}$$

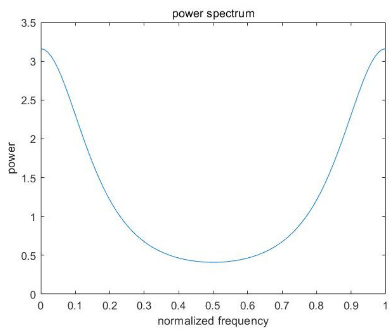
Hence,

$$R_{x_2}(\nu) = \left(\frac{1}{\frac{17}{16}-\frac{1}{2}cos(2\pi\nu)}\right)^2 \tag{18}$$

We plot $R_{x_2}(\nu)$ in MATLAB as Figure 7, which is the power spectrum of signal $x_2(n)$.



(a)$R_{x_1}(\nu)$



(b)$R_{x_2}(\nu)$

Figure 6: The power spectra of signal $x_1(n)$ and $x_2(n)$

D. Task 7

According to formula (7.21) and (7.9) in Collection of Formulas in Signal Processing [1],

$$\mathcal{F}^{-1}\{\frac{1}{\frac{17}{16}-\frac{1}{2}cos(2\pi v)}\} = \frac{16}{15}(\frac{1}{4})^{|k|} \tag{19}$$

$$r_{x_2}(k) = \mathcal{F}^{-1}\{R_{x_2}(v)\} = (\frac{16}{15})^2(\frac{1}{4})^{|k|} * (\frac{1}{4})^{|k|} = (\frac{16}{15})^2 \sum_{l=-\infty}^{\infty}(\frac{1}{4})^{|l|+|k-l|} \tag{20}$$

For $r_{x_2}(k)$ is a even function, we only consider the situation that k > 0, so

$$\begin{aligned} r_{x_2}(k) &= (\frac{16}{15})^2(\sum_{l=-\infty}^{-1}(\frac{1}{4})^{-l+k-l} + \sum_{l=0}^{k}(\frac{1}{4})^{l+k-l} + \sum_{l=k+1}^{\infty}(\frac{1}{4})^{l+l-k}) \\ &= (\frac{16}{15})^2(\frac{1}{4})^k(k+\frac{17}{15}) \end{aligned} \tag{21}$$

When k = 0, $r_{x_2}(0) = \frac{17}{15}$, so

$$r_{x_2}(k) == (\frac{16}{15})^2(\frac{1}{4})^{|k|}(|k|+\frac{17}{15}) \tag{22}$$

We plot $r_{x_2}(k)$ in MATLAB as Figure 7.

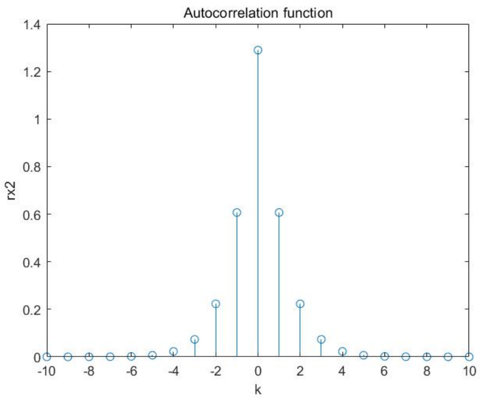


Figure 7: Autocorrelation function of $x_2(n)$

IV. CONCLUSIONS

When discussing Gaussian distribution in section II, we find that the larger the sequence length N, the more closely empirical cdf approximates the original cdf of Gaussian distribution. The larger the correlation coefficient, the closer the shape of the pdf of joint Gaussian distribution is to a line in the XY direction. We also derive formulas such as the conditional distribution.

In section III, two system model with Gaussian noise are analyzed. We derive the periodogram of sinusoidal signals with Gaussian noise and coloured noise. The power spectrum and autocorrelation function of the output of an AR-process and a linear system are also derived.

REFERENCES

[1] Collection of Formulas in Signal Processing, KTH, 2016

[2] https://en.wikipedia.org/wiki/Multivariate_normal_distribution

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