Project Assignment 1
EQ1220 Signal Theory

Hareekeshav Sethumadhavan Srinivasan, Milosh Jankovikj

I. Introduction

The report analyzes the impact of Gaussian distribution on system models through seven tasks split into three parts. First, we examine the accuracy of an estimator based on sequence length affected by Gaussian distribution. Secondly, we focus on adding and subtracting two Gaussian random variables. In addition, we examine the periodogram and noise correlation of two data sequences of a system affected by Gaussian noise. Finally, we are involved in deriving and plotting the PSD and AFC of the input and output signal of an autoregressive (AR) process.

II. Problem Formulation and Solution

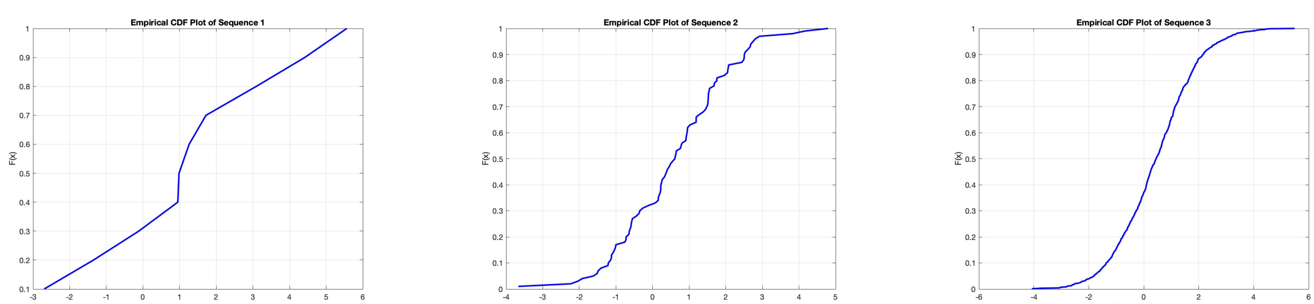
Task 1

We are given three sequences generated according to N(0.5,2). The task aims to demonstrate the impact of sequence length on Gaussian distribution approximation and parameter recovery using the provided data. The estimated means and variances are calculated by $\hat{m} = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{m})^2$.

TABLE I: Task 1 Results

Sequence number	Estimated mean	Estimated variance
1	1.38	6.26
2	0.61	2.19
3	0.44	1.96

The calculations revealed that each sequence has a unique statistical profile with varying mean and variance values. Sequence 1 had the highest values, while Sequence 3 had the lowest. It may not necessarily represent a general phenomenon, as statistical profiles can vary widely based on the characteristics of the data being analyzed. These variances are calculated using the unbiased estimator formula, so they are unbiased estimates of the population variances. Additionally, we have observed an intriguing pattern: when the sequence's length grows, the distribution function tends towards a form of a normal distribution.



(a) $\hat{m}_1 = 1.38$, $\hat{\sigma}^2 = 6.26$ (b) $\hat{m}_1 = 0.61$, $\hat{\sigma}^2 = 2.19$ (c) $\hat{m}_1 = 0.44$, $\hat{\sigma}^2 = 1.96$

Fig. 1: Emperical plots for the given $N(0.5, 2)$.

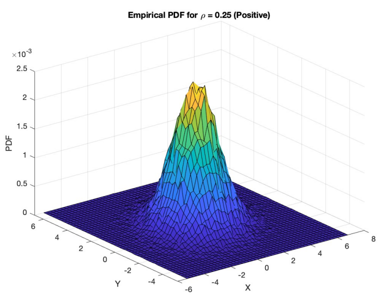
Task 2

We analyze 3D plots based on the joint Gaussian distribution and use the correlation coefficient to determine linear correlation between variables. The distribution is represented by:

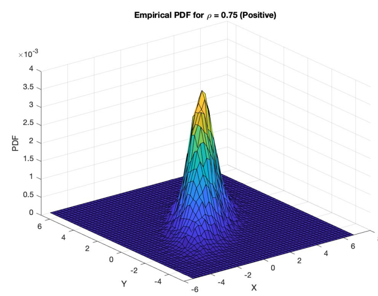
$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma^2} + \frac{(y-\mu_Y)^2}{\sigma^2}\right]\right), \quad (1)$$

where μ_X is the mean of X, μ_Y is the mean for Y, ρ is the correlation coefficient between X and Y and σ^2 is the variance of both X and Y.

Lower correlation coefficients (0.25) produce bulkier 3-D graphs, suggesting weaker variable impact and broader data point spread. Higher coefficients (0.75) result in narrower shapes, indicating stronger linear correlation and more focused data point distribution. Negative correlation coefficients between -0.25 and -0.75 lead to a decrease in one variable when the other variable increases. This causes a flattening of the Gaussian distribution and a downward slope along the antidiagonal in a 2-D plot. Positive correlations cause a 3-D PDF to elongate along the diagonal axis, while negative correlations cause it to elongate along the antidiagonal axis.

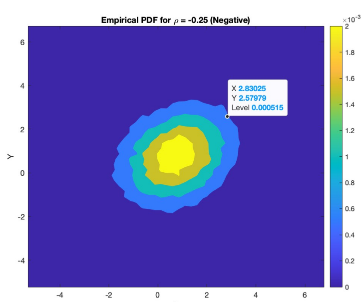


(a) $\rho = 0.25$

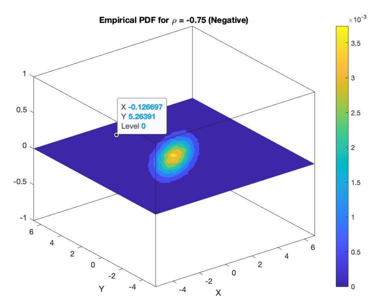


(b) $\rho = 0.75$

Fig. 2: Positive correlation coefficient.



(a) $\rho = -0.25$



(b) $\rho = -0.75$

Fig. 3: Negative correlation coefficient.

Task 3

Firstly, we can determine the probability density function (PDF) of variable Z, which is equivalent to X when Y equals y, using equation (1) and the definition of the conditional distribution:

$$f_Z(z) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_{XY}(z,y)}{f_Y(y)},$$

where $f_Y(y)$ is given by:

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

Hence,

$$\begin{aligned} f_Z(z) &= \\ &= \frac{\sqrt{2\pi\sigma^2}}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(z-\mu_X)^2 - 2\rho(z-\mu_X)(y-\mu_Y) + (y-\mu_Y)^2}{\sigma^2}\right] + \frac{(y-\mu_Y)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(z-\mu_x)^2 - 2\rho(z-\mu_X)(y-\mu_Y)}{\sigma^2}\right]\right) + \frac{(y-\mu_Y)^2}{2\sigma^2}\left(1 - \frac{1}{1-\rho^2}\right). \end{aligned}$$

The mean and variance can be found by addition and subtraction in the numerator: $\sigma_Z^2 = (1-\rho^2)\sigma$ and $\mu_Z = \mu_X + \rho(y - \mu_X)$.

Secondly, suppose $L = X + Y$ and $U = X - Y$. Since a linear combination of Gaussian random variables remains Gaussian, we can obtain the density functions by calculating the mean and variance of $L = X + Y$ and $U = X - Y$:

$$\begin{aligned} \mu_L &= E[L] = E[X] + E[Y] = \mu_X + \mu_Y, \\ \mu_U &= E[U] = E[X] - E[Y] = \mu_X - \mu_Y, \\ \sigma_L &= \sqrt{\sigma_X^2 + \sigma_Y^2}, \\ \sigma_U &= \sqrt{\sigma_X^2 + \sigma_Y^2}. \end{aligned}$$

Following these calculations, one can obtain the *pdfs* for $L = X + Y$ and $U = X - Y$ by substituting the values in (2) for L and U, respectively:

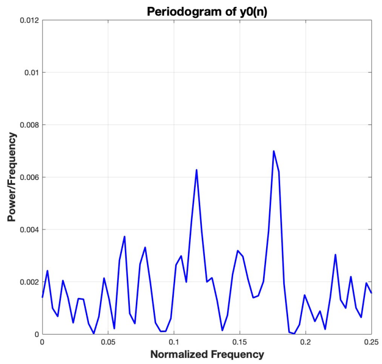
$$\begin{aligned} f_L(l) &= \frac{1}{\sqrt{2\pi}\sqrt[4]{\sigma_X^2 + \sigma_Y^2}} \exp\left(-\frac{(l - \mu_X - \mu_Y)^2}{2\sqrt{\sigma_X^2 + \sigma_Y^2}}\right), \\ f_U(u) &= \frac{1}{\sqrt{2\pi}\sqrt[4]{\sigma_X^2 + \sigma_Y^2}} \exp\left(-\frac{(u + \mu_Y - \mu_X)^2}{2\sqrt{\sigma_X^2 + \sigma_Y^2}}\right). \end{aligned}$$

Task 4

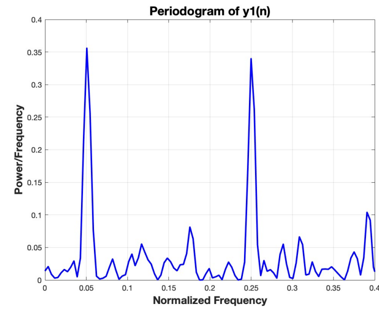
We use Matlab's periodogram to identify signal frequencies and assess the accuracy with interference by comparing recovered frequencies. Accurate frequency estimation is possible without interference in H0. It is important to remember that the noise level can influence frequency estimation accuracy. Higher noise levels can pose challenges in distinguishing underlying signal components. In summary, we detected the following frequencies: 0.391 Hz for H0, 0.054 Hz, 0.252 Hz, and 0.391 Hz for H1, highlighting the accuracy of the recovered sinusoidal frequencies. The plots for the periodogram can be found in figure 4 on the following page.

Task 5

We applied the same methodology to the SinusInNoise2.mat file, which contains a sequence with correlated noise samples. In our analysis of frequency estimation accuracy, we notice that both the periodograms have two distinct peaks and that the periodogram with the colored noise is more spread than the white noise, making it more difficult to estimate the frequencies. The superimposed periodogram can be found in Figure 5 on the following page.



(a) Pure Gaussian WN



(b) Sinusoid mixed with WN

Fig. 4: Periodograms of the two sequences y0(n) and y1(n).

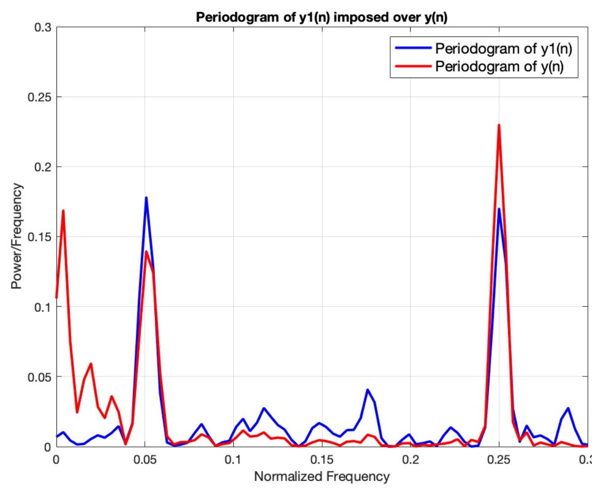


Fig. 5: Periodogram of the signal with white noise (blue) and colored noise (red).

Task 6

We need to derive and plot the power spectra of signals $x_1(n)$ and $x_2(n)$, represented by their autocorrelation functions $R_{x1}(n)$ and $R_{x2}(n)$ respectively. Please refer to the following page for the plots (fig. 6a and 6b).

$$R_{x_1}(v) = \frac{\sigma_x^2}{1-\alpha^2} \sum_{k=-\infty}^{\infty} a^{|k|} e^{-j2\pi vk} = \frac{\sigma_x^2}{1-\alpha^2} \cdot \frac{1-\alpha^2}{1+\alpha^2-2\alpha\cos(2\pi v)} = \frac{1}{1+0.25^2-0.5\cos(2\pi v)}$$

The equation relating the power spectrum $R_{x2}(v)$ to $R_{x1}(v)$ is given by:

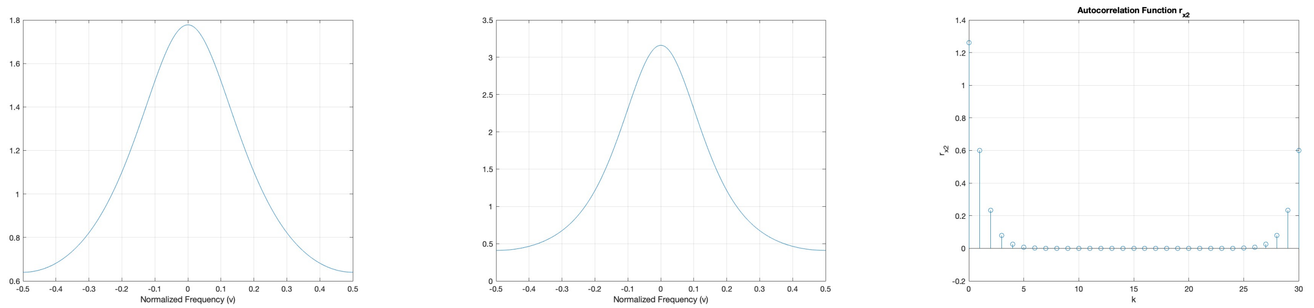
$$R_{x_2}(v) = |H(v)|^2 \cdot R_{x_1}(v),$$

where $H(v)$ is the Fourier transformation of $h_2(n)$:

$$H(v) = \frac{1}{1-0.25e^{-j2\pi v}}.$$

Hence:

$$R_{x_2}(v) = \frac{1}{|1-0.25\cdot e^{-j2\pi v}|^2 \cdot (1+0.25^2-0.5\cdot\cos(2\pi v))}.$$



(a) Power spectrum $R_{x_1}(v)$

(b) Power spectrum $R_{x_2}(v)$

(c) The acf $r_{x_2}(k)$ of the output process $x_2(n)$ $R_{x_1}(v)$

Fig. 6: Power spectrums and autocorrelation function

Task 7

To analyze the output process $x_2(n)$, we need to derive and plot its autocorrelation function, $r_{x2}(k)$. To obtain the acf, we can use the summary of chapter 8 in the textbook [1]. The graph (fig. 6c) displaying the plot for $r_{x2}(k)$ can be found on top of this page.

$$r_{x_2}(k) = \sum_{l=0}^{\infty}\sum_{m=0}^{\infty} h(l)h(m)r_{x_1}(k+m-l) = \sum_{l=0}^{\infty}\sum_{m=0}^{\infty} \beta^{l+m}\alpha^{|k+m-l|}\frac{\sigma_Z^2}{1-\alpha^2} = \frac{1}{1-0.25^2}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty} 0.25^{l+m+|k+m-l|}$$

III. Conclusion

Simply put, the Gaussian distribution is a useful tool for modeling real-world events, and Gaussian noise is commonly used to represent random disruptions in different systems. The inclusion of colored noise makes things more complicated by introducing frequency-dependent characteristics. When an AR(1) process, driven by either Gaussian or colored noise, is sent through an LTI system, the resulting power spectrum and autocorrelation function are affected by the system's frequency response and the input process's characteristics. Therefore, these concepts are incredibly important in signal processing and system analysis.

References

[1] P. Handel, R. Ottoson, H. Hjalmarsson, *Signal Theory*, KTH, 2002
English translation 2007 by Magnus Jansson based on the third edition