Gaussian Distribution Analysis and System Modeling for Signal Detection

Yiyi Miao, Biqi Liu

I. INTRODUCTION

The first part of this report focuses on the correlation properties of the Gaussian distribution. By estimating the mean and variance, it examines the effect of sequence length on the distribution. Additionally, the report explores how different correlation coefficients influence the shape of the probability density function. It provides an in-depth understanding of key concepts such as mean and variance estimation, conditional distributions, and the behavior of sums and differences of correlated variables. Next, we study the transmission of two sinusoidal signals with different normalized frequencies, each affected by white or colored Gaussian noise, and observed the recovery of the signals. The periodogram is utilized to analyze the power spectrum density of these signals. Additionally, we analyze the power spectrum of the autoregressive (AR) process using white noise as the input, as well as the power spectrum and autocorrelation function of the output from a linear time-invariant system, where the AR process served as the input signal.

II. PROBLEM FORMULATION AND SOLUTION

A. Task 1

The three sequences are given and have lengths of 10, 100, and 1000 for s_1, s_2, and s_3, respectively. For each sequence $x_i(n)$, we need to estimate the mean and variance. Given that the sequences are independent and identically distributed (i.i.d.) Gaussian with a known distribution, we will use the following formulas:

$$\mu_i = \frac{1}{N_i}\sum_{n=1}^{N_i} x_i(n) \qquad (1)$$

$$\sigma_i^2 = \frac{1}{N_i}\sum_{n=1}^{N_i}(x_i(n) - \mu_i)^2 \qquad (2)$$

We obtain mean values of s_1, s_2, and s_3 as 1.3829, 0.6135 and 0.4408, respectively. And the variance values of s_1, s_2, and s_3 are 6.2650, 2.1931, and 1.9615, respectively. The CDF describes the cumulative probability of the data distribution. We calculate the empirical CDF for each sequence and plot them on a single graph for comparison. The calculation formula is as follows $\hat{F}(x) = \frac{1}{N}\sum_{i=1}^{N} 1(x \le x_i)$. In Figure 1, we see the empirical cumulative distribution functions (CDFs) of three different data sequences: s_1, s_2, and s_3. The analysis is as follows:

s_1(Blue curve): This is the shortest sequence, and its CDF shows an obvious step-like pattern. This distribution indicates that the sample size is small, and therefore, the CDF changes significantly with the addition of each new data point. This suggests that s_1 does not closely approximate the true Gaussian distribution. s_2(Orange curve): This sequence has more data points than s_1 but fewer than s_3. The CDF is much smoother compared to s_1 and better approximates the theoretical shape of the Gaussian distribution, although some discreteness remains. s_3(Yellow curve): This is the longest sequence. Its CDF curve is very close to the theoretical CDF of the Gaussian distribution, with smooth transitions. The increased data size leads to a more accurate estimate.

Through this task, it is evident that the sequence length has a significant impact on the estimation of the Gaussian distribution. When the data size is small, the estimates are easily influenced by individual samples, leading to large fluctuations. As the data size increases, the distribution estimates gradually become smoother and closer to the true distribution.

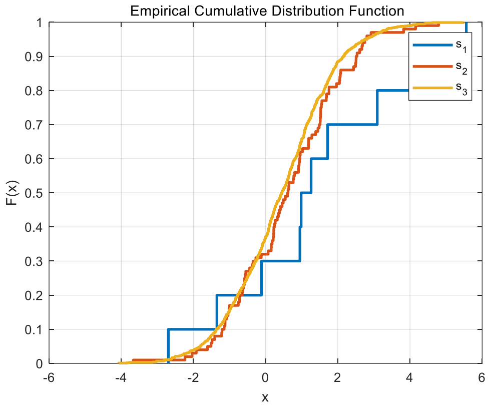


Figure 1
Empirical Cumulative Distribution Function

B. Task 2

According to the task, the general expression for the joint bivariate Gaussian distribution can be simplified to:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} exp\left(-\frac{(x-\mu_x)^2-2\rho(x-\mu_x)(y-\mu_y)+(y-\mu_y)^2}{2\sigma^2(1-\rho^2)}\right) \quad (3)$$

Figure 2 and Figure 3 reflect the joint probability density function (PDF) under different correlation coefficients. We can see that when ρ = 0.25, the distribution is more circular, indicating that X and Y are almost independent. So, the distribution is more symmetric. When $\rho = 0.75$, joint distribution becomes elongated along the diagonal, reflecting a stronger positive correlation. This indicates that as one variable increases, the other also tends to increase.

If the pdf is characterized from $-\rho$ instead of ρ, the shape will be elongated along the negative diagonal (from top-left to bottom-right). It reflects that X and Y are negatively correlated.

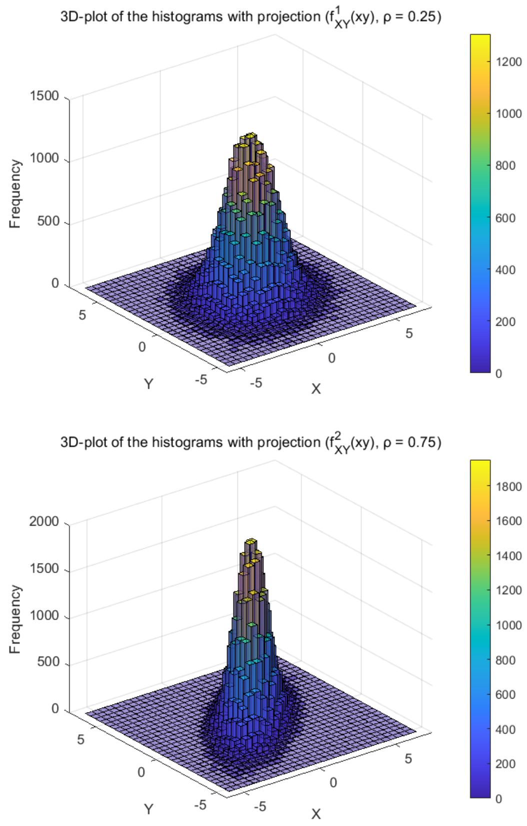
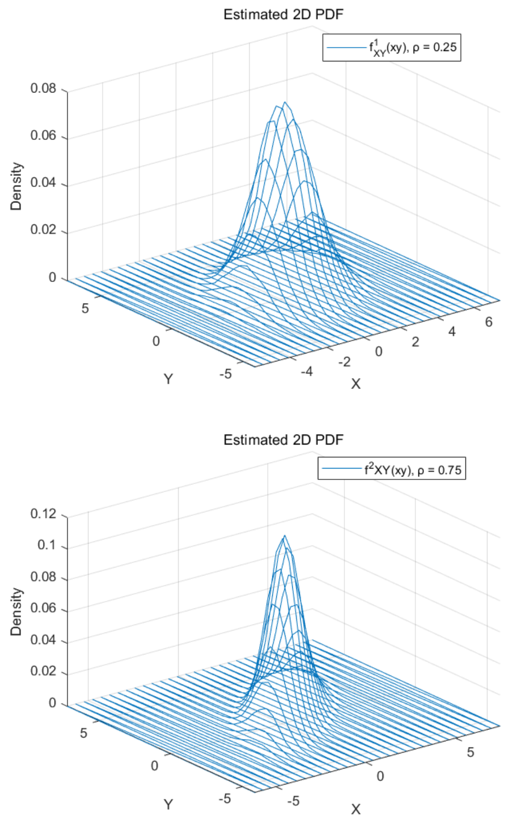


Figure 2 3D-plot of the Histograms



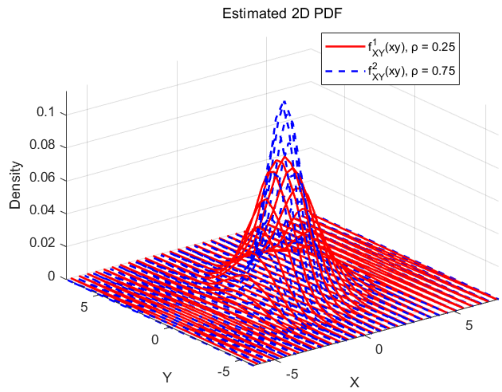


Figure 3 Estimated 2D PDF

C. Task 3

The marginal distribution $f_Y(y)$ can be found by integrating out x from the joint distribution, which gives:

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(y-\mu_y)^2}{2\sigma^2}) \qquad (4)$$

Now, dividing the joint distribution by the marginal $f_Y(y)$, we get the conditional distribution:

$$f_{X|Y=y}(x, y) = \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} exp\left(-\frac{(x-\rho y-\mu_x+\rho\mu_y)^2}{2\sigma^2(1-\rho^2)}\right) \quad (5)$$

The conditional pdf for the random variable $Z = X \mid Y = y$. This is equivalent to the conditional distribution derived above:

$$f_z(z) = \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} exp\left(-\frac{(z-\rho y-\mu_x+\rho\mu_y)^2}{2\sigma^2(1-\rho^2)}\right) \qquad (6)$$

For two normally distributed random variables X and Y, the sum $W = X + Y$ follows another normal distribution. We calculate its mean and variance as follows:

$$E[W] = E[X] + E[Y] = \mu_x + \mu_y \qquad (7)$$

$$Var(W) = Var(X) + Var(Y) + 2Cov(X, Y) \ (8)$$

Since $Var(X) = Var(Y) = \sigma^2$ and $Cov(X, Y) = \rho\sigma^2$, the variance of W:

$$Var(W) = \sigma^2 + \sigma^2 + 2\rho\sigma^2 \qquad (9)$$

Thus, the pdf of W is a Gaussian distribution with mean 0 and variance $2\sigma^2(1 + \rho)$

$$f_W(w) = \frac{1}{\sqrt{4\pi\sigma^2(1+\rho)}} exp(-\frac{(w-(\mu_x+\mu_y))^2}{4\sigma^2(1+\rho)}) \quad (10)$$

Similarly, for the difference V = X - Y, we can derive the pdf of V is a Gaussian distribution with mean 0 and variance $2\sigma^2(1 - \rho)$:

$$f_V(v) = \frac{1}{\sqrt{4\pi\sigma^2(1-\rho)}} exp(-\frac{(v-(\mu_x-\mu_y))^2}{4\sigma^2(1-\rho)}) \qquad (11)$$

D. Task 4

Gaussian white noise has been widely used to model the noise during transmission. In Task 4, we study the impact of white noise on sinusoidal signal recovery. We load the given signals and plot the periodogram of the output sequences for H0 and H1. It is clear to see that there are two peaks at $\nu = 0.05, 0.25$ in the blue curve. As a result, the red curve corresponds to H0 while the blue curve corresponds to H1. Although white Gaussian noise slightly broadens the peaks in the power spectral density plot, the sinusoidal frequencies still stand out prominently.

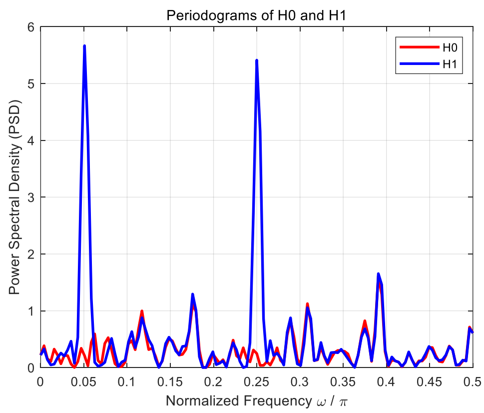


Figure 4 Periodograms of H0 and H1

E. Task 5

In some cases, we might have colored i.e. correlated, noise in the channel to corrupt the signals. In Task 5, we focus on the impact of colored noise on sinusoidal signal recovery. Similarly, we load the given data and plot the periodogram. In this case, there are two peaks at different magnitudes around the expected frequencies: $\nu = 0.05, 0.25$. Compared with Task 4, $\nu_0 = 0.05$ may experience more noise interference, lowering its peak, while $\nu_1 = 0.25$ is less affected with a higher peak. Unlike white noise, colored noise adds a frequency-dependent distortion to the signal. Hence, the peaks of different frequencies with the same initial magnitude can be distorted.

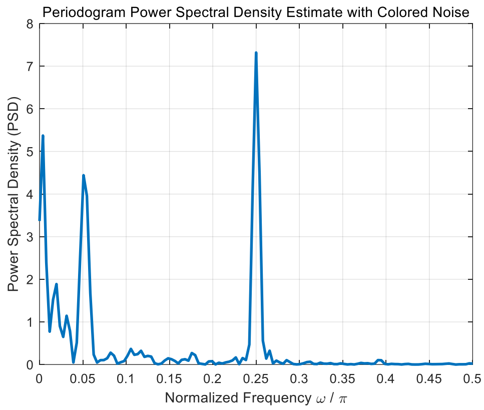


Figure 5 Periodogram of Colored Noise

F. Task 6

Last but not least, we focus on the AR(1) process $x_1(n)$, driven by a white noise $z(n)$ with variance σ_z^2. Another linear system with an impulse response, $h_2(n)$, receives the output sequence $x_1(n)$ as input, and the corresponding output signal is denoted as $x_2(n)$. The spectra of $x_1(n)$ and $x_2(n)$ are studied in Task 6. Additionally, we look into the autocorrelation function of $x_2(n)$ in Task 7. Specifically, the AR(1) process $x_1(n)$ is modeled as $x_1(n) = \alpha x_1(n - 1) + z(n)$. The impulse response of the linear system is $h_2(n) = \beta^n u(n)$, where $u(n)$ is the unit step function. The parameters are set as: $\sigma_z^2 = 1$, $\alpha = 0.25, \beta = 0.25$. Hence, the output $x_2(n)$ when system stimulated by $x_1(n)$ can be written as:

$$x_2(n) = \sum_{k=0}^{+\infty} h_2(k) x_1(n-k) \tag{12}$$

According to (8.17) and (8.21) from the textbook[1], the power spectrum of AR(1) process $x_1(n)$ is given by the magnitude-squared of its frequency response multiplied by the variance of the white-noise input, as in Eq. (13). Likewise, the power spectrum of $x_2(n)$ is the product of the power spectrum density function of $x_2(n)$ and the magnitude-squared of the frequency response, as in Eq. (14).

$$\begin{aligned} R_{X_1}(\nu) &= \left|H_1(\nu)\right|^2 \sigma_z{}^2 \\ &= \left|\frac{1}{1-\alpha e^{-j2\pi\nu}}\right|^2 \sigma_z{}^2 \\ &= \frac{1}{1.0625 - 0.5\cos(2\pi\nu)} \end{aligned} \tag{13}$$

$$\begin{aligned} R_{X_2}(\nu) &= R_{X_1}(\nu) \cdot \left|H_2(\nu)\right|^2 \\ &= \left|\frac{1}{1-\alpha e^{-j2\pi\nu}}\right|^2 \sigma_z{}^2 \cdot \left|\frac{1}{1-\beta e^{-j2\pi\nu}}\right|^2 \\ &= \frac{1}{\left(1.0625 - 0.5\cos(2\pi\nu)\right)^2} \end{aligned} \tag{14}$$

Plotting these two spectra in the same figure, we can see from Figure 6 that the two spectra have similar low-pass behavior, meaning most of the signal's power is concentrated at lower frequencies. Also, across all frequencies, the power spectral density of $x_2(n)$ is approximately twice that of $x_1(n)$, which corresponds to the analytical expressions. Comparing the two curves, the blue curve representing $x_2(n)$ has a sharper decline at higher frequencies, which is expected because filtering $x_1(n)$ through $h_2(n)$ further suppresses high-frequency components. As a result, compared to $x_1(n)$, the low-frequency components of $x_2(n)$ are enhanced.

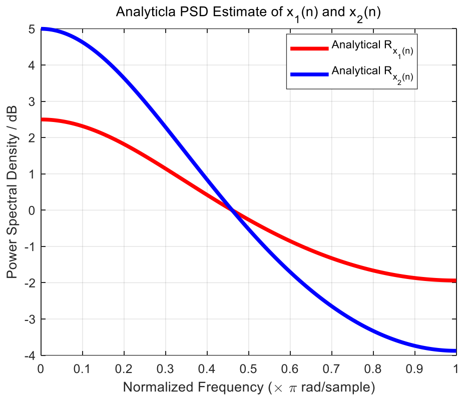


Figure 6
Analytical PSD Estimate of $x_1(n)$ and $x_2(n)$

G. Task 7

The autocorrelation function of the AR(1) process $x_1(n)$, denoted as $r_{x_1}(k)$, satisfies:

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

From (8.16) in the textbook [1], we can derive the autocorrelation function of $x_2(n)$, denoted as $r_{x_2}(k)$, in Eq. (16).

$$\begin{aligned} r_{x_2}(k) &= \sum_{l=-\infty}^{\infty}\sum_{p=-\infty}^{\infty} h(l)h(p)r_x(k+p-l) \\ &= \sum_{l=-\infty}^{\infty}\sum_{p=-\infty}^{\infty} \beta^l u(l)\beta^p u(p)\alpha^{|k+p-l|}\frac{\sigma_z^{\ 2}}{1-\alpha^2} \\ &= \sum_{l=0}^{\infty}\sum_{p=0}^{\infty} 0.25^{p+l}\cdot 0.25^{|k+p-l|}\cdot\frac{1}{1-0.25^2} \\ &= \frac{16}{15}\sum_{l=0}^{\infty}\sum_{p=0}^{\infty} 0.25^{p+l+|k+p-l|} \end{aligned} \tag{16}$$

In Figure 7, the autocorrelation function $r_{x2}(k)$ achieves its maximum value when the lag is zero and shows a sharp decline after the first few lags, which indicates that the signal $x_2(n)$ has a limited correlation between distant values in time.

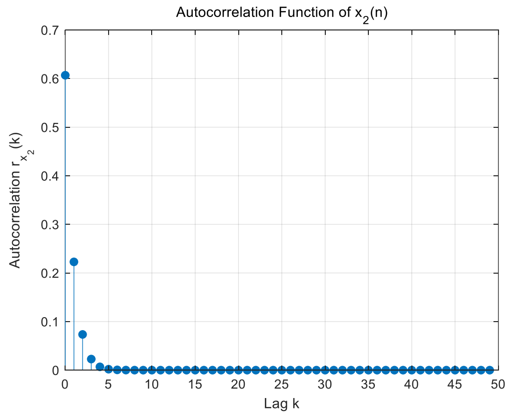


Figure 7 Autocorrelation Function of $x_2(n)$

III. CONCLUSIONS

This study emphasizes the role of sequence length, correlation, and distribution analysis in understanding Gaussian processes. Task 1 showed that longer data sequences improve the accuracy of Gaussian distribution estimates. Task 2 demonstrated that higher correlation coefficients create more elongated joint distributions. Task 3 explored how correlation influences the behavior of sums and differences in correlated variables. In Task 4, white noise kept the main frequencies of sinusoidal signals clear, while Task 5 showed that colored noise distorted the spectral content, making frequency identification harder. In Task 6, the power spectra of the AR(1) process and its filtered version $x_2(n)$ showed low-pass behavior, with $x_2(n)$ further attenuating high-frequency components. Finally, Task 7 analyzed the autocorrelation function of $x_2(n)$. The results displayed a rapid decay after the first few lags, indicating limited correlation over time. The study highlights the challenges noise poses in signal processing and suggests further exploration of different noise types on signals and systems.

REFERENCES

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