Yann DEBAIN

Project Assignment 1

I. INTRODUCTION

This project aims to understand the fundamental concepts in probability theory and signal processing. All these concepts will be illustrated with Gaussian distribution because of its importance in system models in signal processing.

This project will tackle the concept of parameter estimator, joint and conditional distribution and system models with different Gaussian noises (white or colored noise).

These concepts will be illustrated through mathematical formula and visual representation implemented with MATLAB.

II. THE GAUSSIAN DISTRIBUTION

The aim of this section is to interpret the impact of the parameters on the Gaussian distribution and to present some properties of the Gaussian distribution.

Task 1

In this task, we want to illustrate the impact of the sequence length in the approximation of a Gaussian distribution and the recovery of its parameters.

We assume that the mean and the variance of these sequences are unknown, we want to estimate them through unbiased parameter estimators.

$$Mean \quad estimator : \hat{M}_x = \frac{1}{N_i} \sum_{n=1}^{N_i} X_i(n) \tag{1}$$

$$Variance \quad estimator : \hat{S}_x^2 = \frac{1}{N_i - 1} \sum_{n=1}^{N_i} (X_i(n) - \hat{M}_x)^2 \tag{2}$$

We implement this formula in MATLAB (through the functions mean(.) and var(.)) and we have the result below :
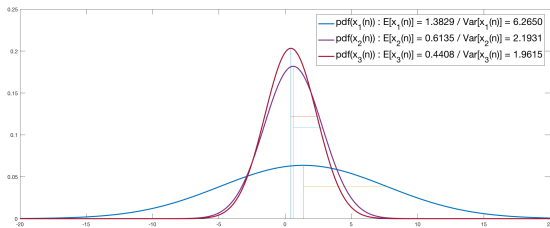

FIGURE 1: Visual representation of the three sequences $\{x_i(n)\}_{i=1}^3$ and estimation of their parameters

We can observe that the bigger N_i is, better the approximation of the Gaussian parameters is.

Thanks to the unbiased estimators, we know that we have, on average, the proper result but this is the fact the estimators (1) and (2) are consistant that explain this point $\begin{cases} Var[\hat{M}_x^2] = \frac{\sigma_x^2}{N_i} \\ Var[\hat{S}_x^2] = 2\frac{\sigma_x^4}{N_i - 1} \end{cases}$.

It is possible to approximate a Gaussian distribution as well as its parameters, with a degree of precision, with a long sequence.

Task 2

Giving X and Y, two random variables with $X \hookrightarrow N(m_X, \sigma_X^2 = \sigma^2)$ and $Y \hookrightarrow N(m_Y, \sigma_Y^2 = \sigma^2)$

The joint Gaussian distribution is :

$$f_{XY}^{(i)}(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho_i^2}}e^{-\frac{(x_i-m_X)^2 - 2\rho_i(x_i-m_X)(y_i-m_Y)+(y_i-m_Y)^2}{2\sigma^2(1-\rho_i^2)}} \tag{3}$$

We want to study the impact of the correlation coefficient in the shape of the joint distribution. First we will visualize the shape of the joint distribution for two given correlation coefficient, then make a generalization.

For two sequences $\{x_i(n), y_i(n)\}_{i=1}^2$, the correlation coefficient ρ_i between X and Y takes one of the values $\{0.25, 0.75\}$. We visualize the shape of the joint convolution for each sequence.
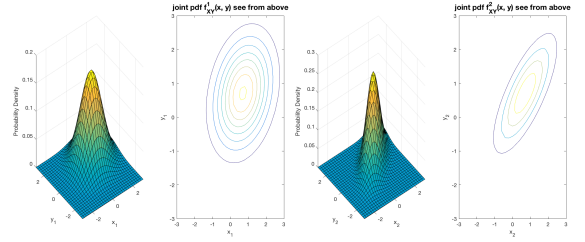

FIGURE 2: 3D representation of the two pdfs and seen from above

To determine which correlation coefficient corresponds to which shape, we have to know that the correlation coefficient represents the relationship between the variable. In this case, the rounder the shape is, the weaker $|\rho_i|$ is ; and the more linear the shape is and the stronger $|\rho_i|$ is.

So we can determine from Figure 2 that $\begin{cases} \rho = 0.25 \Rightarrow \{x_1(n), y_1(n)\} \\ \rho = 0.75 \Rightarrow \{x_2(n), y_2(n)\} \end{cases}$

Changing the sign of the correlation coefficient ρ_i symmetrically inverts the shape of the distribution of the joint with respect to the mean of a variable.

Task 3

Using (3), we have $f_{XY}(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}}e^{-\frac{(x-m_X)^2 - 2\rho(x-m_X)(y-m_Y)+(y-m_Y)^2}{2\sigma^2(1-\rho^2)}}$

We want to determine some properties of the Gaussian distribution. We assume a correlation coefficient ρ, between X and Y , in all cases.

— $f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{\sigma\sqrt{1-\rho^2}\sqrt{2\pi}}e^{-\frac{[x-(m_X+\rho y - \rho m_Y)]^2}{2\sigma^2(1-\rho^2)}}$

— Giving X and Y, two random variables with $X \hookrightarrow N(m_X, \sigma_X^2 = \sigma^2)$ and $Y \hookrightarrow N(m_Y, \sigma_Y^2 = \sigma^2)$

The sum of two Gaussian is also Gaussian
$E[X+Y] = E[X] + E[Y] = m_X + m_Y$
$Var[X+Y] = Var[X] + Var[Y] + 2Cov(X,Y) = 2\sigma^2(1+\rho)$

We have $X + Y \hookrightarrow N(m_X + m_Y, 2\sigma^2(1+\rho))$

$$f_{X+Y}(k) = \frac{1}{\sqrt{4\pi\sigma^2(1+\rho)}}e^{-\frac{(k-m_X-m_Y)^2}{4\sigma^2(1+\rho)}}$$

— Giving X and Y, two random variables with $X \hookrightarrow N(m_X, \sigma_X^2 = \sigma^2)$ and $Y \hookrightarrow N(m_Y, \sigma_Y^2 = \sigma^2)$

The sum of two Gaussian is also Gaussian
$E[X - Y] = E[X] - E[Y] = m_X - m_Y$
$Var[X - Y] = Var[X] + Var[Y] - 2Cov(X,Y) = 2\sigma^2(1 - \rho)$

We have $X - Y \hookrightarrow N(m_X - m_Y, 2\sigma^2(1 - \rho))$
$$f_{X-Y}(k) = \frac{1}{\sqrt{4\pi\sigma^2(1-\rho)}}e^{-\frac{(k-m_X+m_Y)^2}{4\sigma^2(1-\rho)}}$$

Conclusion

We have seen in this section that to estimate the parameter of a Gaussian distribution, the longer the sequence is, the better the estimation is. We have also studied some properties of the joint Gaussian distribution, in particular, the impact of the correlation coefficient which shape of the joint distribution.

III. SYSTEM MODELS WITH GAUSSIAN NOISE

This section presents different models of Gaussian noise systems. We will interpret the impact of noise on the estimation of the parameters of the desired signal.

Task 4

In this task, we want to transfer two sinusoïdal signals but the transmission is disturbed by a Gaussian noise.

We have two possible outcomes, H0 : $y_0(n) = w(n)$ and H1 : $y_1(n) = \sum_{k=0}^{1} sin(2\pi\nu_k n + \phi_k) + w(n)$. To determine which output sequence contains the sinusoïdal information, we estimate the power spectrum through a technic called periodogram and we get the results below :

The two sinusoidal signals have normalized (with the sampling frequency fs) frequencies $\nu_0 = 0.05$ and $\nu_1 = 0.25$ and initial phases φ_0 and φ_1, respectively. So the power spectrum of the output with the sinusoïdal information must contains power peaks on the normalized frequencies ν_0 and ν_1.
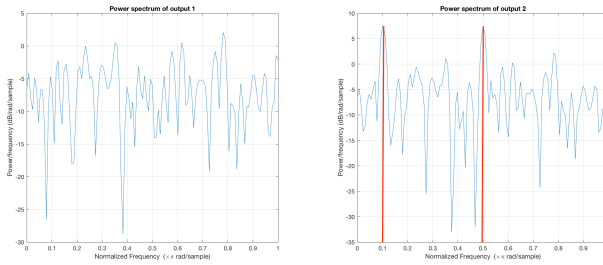
FIGURE 3: Periodogram of the different outputs

On the power spectrum of output 2, we can see that there is two power peaks on the normalized frequencies $2 * \nu_0 = 0,1$ and $2 * \nu_1 = 0,5$. (We have $2 * \nu_0 = 0,1$ and $2 * \nu_1 = 0,5$ because on the periodogram the frequencies are normalized by π rad/sample instead of 2π rad/sample).

So we have, the power spectrum of output 1 is the power spectrum of H0 and the power spectrum of output 2 is the power spectrum of H1.

For a white noise, the estimation accuracy of the frequencies depends on the signal noise ratio (SNR). If the power of the noise is stronger than the power of the signal, it will be difficult to estimate the frequencies. It is therefore necessary to keep a large SNR to find the required information in a periodogram.

Task 5

Sometimes, the transmission could be disturbed by a colored noise. We want to analyze the power spectrum to see the impact of the noise on the estimated power spectrum.
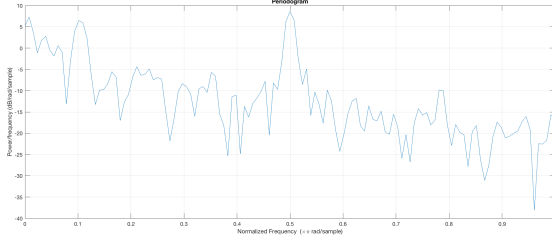

FIGURE 4: Periodogram

Most of the noise is situated in the low frequencies area so it is difficult to distinguish the low frequency $2 * \nu_0 = 0,1$ of the sinusoïdal signal. On the other hand we can properly distinguish the frequency at $2 * \nu_1 = 0,5$ because the noise power is weak in this frequency range.

For colored noise, the estimation accuracy of the frequencies is difficult if the frequencies are situated in the low frequencies area.

Task 6

$x_1(n) = \alpha x_1(n-1) + z(n)$
$x_2(n) = \sum_{k=0}^{+\infty} h_2(n)x_1(n-k)$
We want to determine the power spectra of the signals $x_1(n)$ and $x_2(n)$.

$$R_x(\nu) = F[r_x(k)] \tag{4}$$

For a linear time invariant system we have :

$$R_{output}(\nu) = |H(\nu)|^2 R_{input}(\nu) \tag{5}$$

To determine the power spectrum we will use the formula (4) and (5).

— $x_1(n) = \alpha x_1(n-1) + z(n)$

$\begin{cases} \sigma_{x_1}^2 = \frac{\sigma_z^2}{1-\alpha^2} \\ r_{x_1}(k) = \alpha^{|k|} r_{x_1}(0) = \alpha^{|k|} \sigma_{x_1}^2 \end{cases} \Leftrightarrow r_{x_1}(k) = \alpha^{|k|} \frac{\sigma_z^2}{1-\alpha^2}$

We deduce that $R_{x_1}(\nu) = F[r_{x_1}(k)] = \frac{\sigma_z^2}{1+\alpha^2-2\alpha cos(2\pi\nu)} = \frac{1}{\frac{17}{16}-\frac{1}{2}cos(2\pi\nu)}$

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

We have a linear time invariant system so we can use the formula (5) and the formula (7.19) in [1] :
$H(\nu) = F[h_2(n)] = \frac{1}{1-\beta e^{-j2\pi\nu}}$

We deduce that $R_{x_2}(\nu) = |H(\nu)|^2 R_{x_1}(\nu) = \frac{\sigma_z^2}{(1-2\beta cos(2\pi\nu)+\beta^2)(1-2\alpha cos(2\pi\nu)+\alpha^2)} = \frac{1}{(\frac{17}{16}-\frac{1}{2}cos(2\pi\nu))^2}$
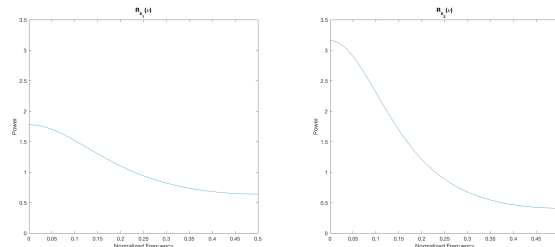

FIGURE 5: Power spectra of $x_1(n)$ and $x_2(n)$

Task 7

In the last task, we derive $R_{x_2}(\nu)$ without knowing $r_{x_2}(k)$. Now we want to determine $r_{x_2}(k)$. For that we will use the inverse Fourier transform of the power spectrum.

$$r_{x_2}(k) = F^{-1}[R_{x_2}(\nu)] \tag{6}$$

We obtain $r_{x_2}(k) = \sigma_z^2 \frac{1}{(1-\beta^2)(1-\alpha^2)} \beta^{|k|} * \alpha^{|k|}$

$r_{x_2}(k) = \frac{256}{225}(0.25)^{|k|} * 0.25^{|k|}$

$r_{x_2}(k) = \frac{256}{225} \frac{0.25^{|k|}(|k|+1)+0.25^{|k|+2}(|k|-1)}{1-0.25^2}$
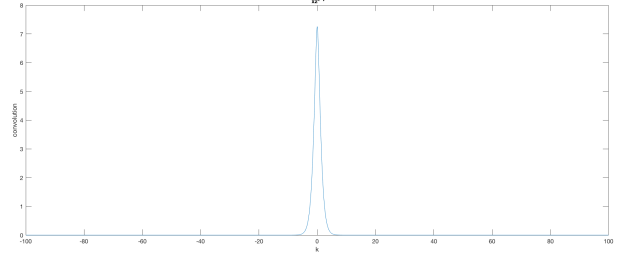
$r_{x_2}(k) \approx 0.25^{|k|}(|k|+1)$

FIGURE 6: ACF of $x_2(n)$

Conclusion

This section has allowed us to understand how to estimate the required information in a system model with a Gaussian noise. In a power study, it is important to have a large SNR in order to recover the information in the power spectrum.

IV. CONCLUSIONS

In this project, we have seen some properties of a Gaussian distribution and the impact of its parameters. We have also seen how to estimate the required information in the presence of a Gaussian noise (white or colored noise).

The SNR, sequence length and correlation coefficient are important information to know because a poor knowledge of this information can lead to a bad study with misinterpretations.

These concepts are important because of system modeling in electronics and signal processing which often uses a Gaussian noise to model the noise present.

V. REFERENCES

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