KTH Electrical Engineering

Project Assignment 1
EQ1220 Signal Theory

Stockholm 2021

School of Electrical Engineering and Computer Science
Royal Institute of Technology

Introduction

Fundamental concepts in elementary probability theory and signal processing are essential towards the solid understanding of more complicated principles in engineering. In this project we shall go through some simple, yet important, concepts in order to obtain "hands-on" experience of the material taught in class. We are going to experiment mainly with the Gaussian distribution, which is commonly adopted in various system models in engineering applications. We shall further examine concepts related to parameter estimation and other signal properties, involving both mathematical derivations and visual representations in order to be able to formulate rigorous conclusions.

The goal of this project is twofold. From the technical viewpoint, the student is expected to be able to apply simple principles discussed in class and carry out MATLAB implementations. From a presentation viewpoint, the student is expected to be able to present the problem and the solution in a self-contained, cohesive and comprehensible report such that any peer reader can understand and reproduce the results. Both aspects are **equally important** towards the successful completion of this project.

Project Tasks

This project is partitioned into different tasks, which span the material of the first chapters of the course textbook. The series of the tasks, however, does not follow strictly the order in which they have been taught. Nevertheless they have been formulated and sorted in a way that solution and successful completion of one task assists the solution of the next one.

The Gaussian Distribution

The file **Gaussian1D.mat** contains three sequences $\{x_i(n)\}_{i=1}^{3}$, of different lengths N_i, whose elements are independent and identically distributed and they have been generated according to $\mathcal{N}(0.5, 2)$. The purpose of this task is to illustrate the impact of the sequence length in the approximation of a Gaussian distribution from given data as well as on the recovery of its parameters.

Task 1: *Estimate the mean and the variance of the distribution, from each sequence $\{x_i(n)\}_{i=1}^{3}$. Plot the empirical distributions and comment on the impact of the sequence length N_i on the approximation of the original distribution from the given data.*

Hint: Typical estimators for the mean and variance from a continuous data sequence are given by (2.38), (2.39) *in the textbook. The corresponding estimators for the discrete model are found by replacing integration with summation and the time index t with the sample index n. You can also check section* 9.3.

Next we shall use the file **Gaussian2D.mat** which contains two matrices, each containing a data sequence $\{(x_i(n), y_i(n))\}_{i=1}^{2}$, whose elements are again i.i.d Gaussian and which have been generated from bivariate Gaussian distributions with pdf $f_{XY}^{(i)}(x, y)$. The random variables X and Y have equal variances $\sigma_x^2 = \sigma_y^2 = \sigma^2$. Moreover the correlation coefficient ρ_i between X and Y takes one of the values $\{0.25, 0.75\}$.

Task 2: *Write down the general expression for the joint Gaussian distribution $f_{XY}^{(i)}(x, y)$. Provide a 3D-plot of the empirical pdfs using the given data and match each plot to the possible values for ρ_i. Comment on the impact of the correlation coefficient on the shape of the pdf. What kind of shape/behavior would you expect if the pdf was characterized from $-\rho_i$, instead of ρ_i?*

Task 3: *Using the general formula, derived in the previous task, for $f_{XY}(x, y)$ and the definition of the conditional distribution $f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)}$ find in closed-form the conditional pdf $f_Z(z)$ of the random variable $Z = X|Y = y$. Also derive the pdfs for $X + Y$ and $X - Y$. Assume a correlation coefficient ρ, between X and Y, in all cases.*

System Models with Gaussian Noise

The previous subsection has given a flavor of simple properties of the Gaussian distribution and also some pictorial representations. In this part we shall go one step further and look into a little more complicated system models.

Assume that we want to transmit two sinusoidal signals with normalized (with the sampling requency f_s) frequencies $v_0 = 0.05$ and $v_1 = 0.25$ and initial phases ϕ_0 and ϕ_1, respectively. Typically the transmission is corrupted by Gaussian noise, which has certain statistical properties. The file **SinusInNoise1.mat** contains data sequences, each with a different signal representation. The possible outcomes are given below as

$$\begin{aligned} \text{H0}: &\quad y_0(n) = w(n) \\ \text{H1}: &\quad y_1(n) = \sum_{k=0}^{1} sin(2\pi v_k n + \phi_k) + w(n) \end{aligned} \tag{1}$$

where $w(n)$ is a white noise sequence with variance σ_w^2 and the signal-to-noise ratio (SNR) of each sinusoid, defined as $\text{SNR}_k = \frac{A_k^2}{\sigma_w^2}$, is 0dB. In order to examine which case holds for a given observation sequence we carry out some simple tests.[1]

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.*

In some cases we might have non-white (colored) noise, disturbing our transmission. The file **SinusInNoise2.mat** contains a sequence with correlated noise samples.

Task 5: *Using the periodogram, comment on the impact of noise correlation on the estimation accuracy of the normalized frequencies v_0 and v_1, respectively.*

We shall now detour and look into another simple system model, the AR(k) process, that we have seen in class. In this case, a white noise sequence $z(n)$ with variance σ_z^2 is the input that drives an AR(1) process $x_1(n)$, modeled as

$$x_1(n) = \alpha x_1(n-1) + z(n) \tag{2}$$

The output sequence $x_1(n)$ is fed as input to another linear system with impulse response $h_2(n) = \beta^n u(n)$, where $u(n)$ is the unit step function. The output $x_2(n)$ can be written as

$$x_2(n) = \sum_{k=0}^{+\infty} h_2(k) x_1(n-k) \tag{3}$$

Assume values $\sigma_z^2 = 1, \alpha = 0.25$ and $\beta = 0.25$ for the system parameters.

[1]In some fields this is called *hypothesis testing.*

Task 6: *Derive and plot the power spectra $R_{x_1}(v)$ and $R_{x_2}(v)$ of the signals $x_1(n)$ and $x_2(n)$, respectively.*

Task 7: *Derive and plot the acf $r_{x_2}(k)$ of the output process $x_2(n)$.*

Written Report

The assignment should be presented in a written report (3-5 pages single or double column). The written report should be organized as is done with technical publications and should be written in English. The template for reports (latex and MS word) are available in the uploaded zip file.

In general, a technical report consists of a brief introduction, where the problem under consideration is briefly described, the main part where the problem is formulated and where the solution is provided, and the conclusions section, where one reflects on the output. It is suggested to formulate your report such that it follows the structure of this assignment, i.e., the section division, and such that it addresses the assigned tasks sequentially.

Do not just state the answers but justify and frame them using theory (You can even reproduce on paper what is written in the hints, if you make use of them in the solution, **but not verbatim**!). Explain how each task was solved (especially the mathematical principles) and what the results were. If you have many derivations place them in an appendix at the end of the report. The MATLAB code is not necessary at all. The report should be written such that a fellow student can understand and reproduce your results without having access to this instruction. Support your conclusions with either numerical results and/or graphical results (the latter should be almost clear by just looking at the figure). In case you are using sophisticated MATLAB functions in order to solve any exercise provide a short explanation in your own words of what the function is doing.

The report will be given the grade **Pass** or **Fail**. The grading includes clarity of presentation, content and layout. If the report gets the grade Fail, it can be revised and submitted a second time. If the revised report also fails, you will have to redo the project next year. The assignment should be performed individually or in groups of two students (recommended!). It is allowed to discuss (orally, without pen and paper) the problem formulation, and different methods of solving it, with others but you are not allowed to share calculations, Matlab code or images! The report should be the product of individual group work and it is not allowed to reproduce or rephrase other people's work. This is considered as plagiarism and KTH has very strict rules against it. If you use MS Word provide the equations using the embedded equation editor or MathType (found at `https://www.kth.se/en/student/kth-it-support/software/download`). Handwritten reports are not acceptable!

Please submit the report via Canvas.