ECE499/CS499 Statistical Signal Processing (Fall 2021)

4 credits, MW 12:00-01:50pm, Bexell Hall 322

1. **Instructor**: Prof. Jinsub Kim (jinsub.kim@oregonstate.edu), 3011 Kelley Engineering Center
2. **Prerequisite**:
   a. Probability & Random Process (ECE 353) **or** Introduction to Statistics for Engineers (ST 314)
   b. Familiarity with Python **or** MATLAB (you will be asked to implement some algorithms and perform experiments using either Python or MATLAB; you can choose)

3. **Textbooks**: We will use the following two resources, which are *available online for free.*
   a. Lecture notes by A. V. Oppenheim and G. C. Verghese, *MIT EECS 6.011 Introduction to Communication, Control, and Signal Processing (Spring 2010) (Available through MIT OpenCourseWare)*
   b. G. James, D Witten, T. Hastie, and R. Tibshirani, "*An Introduction to Statistical Learning with Applications in R*", 2nd Ed., Springer, 2021 (PDF available at https://www.statlearning.com)

4. **Target audience**:
   a. *Engineering students* who want to learn fundamentals of detection, estimation, and statistical learning as well as how they are applied to many engineering problems
   b. *Computer Science students* (interested in machine learning) who want to learn about classical detection and estimation theory and statistical perspective of machine learning (note that most machine learning problems are essentially statistical inference problems!)

5. **Course description:**

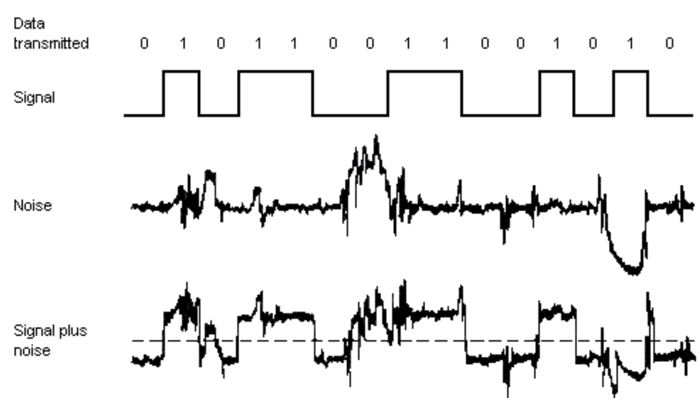

Figure 1. Transmitted signal (top), random channel noise (middle), and random noisy signal at the receiver (bottom) due to the additive channel noise.

Statistical inference problems exist in almost every engineering discipline. For instance, in digital communications, the receiver needs to *detect* a sequence of bits sent by the transmitter based on the received signal which is corrupted by random noise (see the right figure). In controlling an unmanned aerial vehicle (UAV), a localization technique is needed to *estimate* UAV's location and speed based on various sensor measurements. A RADAR system needs to be equipped with *detection* and *classification* algorithms that can be used to detect and classify objects within the monitored area based on noisy RADAR measurements. All these problems essentially boil down to the following question:

"Given an observation of random variable X (e.g., UAV sensor data), how can we estimate the value of an unknown Y (e.g., UAV's location) accurately?"

At the successful completion of this course, students will be able to answer the above question.

6. Course topics:

The course consists of two parts. In **Part 1 (Detection and Estimation)**, assuming that the probability distribution of X and Y is known, we will discuss optimal detection and estimation rules that we can use to draw an inference about Y based on observation of X optimally with respect to some popular criteria. The topics in Part 1 include Bayesian detection (MAP detector), Non-Bayesian detection (Likelihood ratio test and Neyman-Pearson Lemma), Bayesian estimation (MMSE estimator and linear MMSE estimator), Non-Bayesian estimation (Maximum likelihood estimator), and various applications in engineering.

In **Part 2 (Statistical Learning)**, we will discuss the same question, but in the setting where the probability distribution of X and Y is *unknown*; instead, we are given some samples from the distribution (the so-called *training dataset*), from which we can learn about the statistical dependency between X and Y. This is a typical supervised learning formulation considered in machine learning and has been popularly employed for many engineering problems due to the difficulty of coming up with an accurate probability model for complex data. In this course, we will put emphasis on the *statistical perspective* of machine learning; the main discussion will be about how the ideas of optimal detection and estimation (learned in Part 1) are adapted for solving statistical learning problems. The topics in Part 2 include the bias-variance tradeoff, parametric approaches (linear regression, logistics regression, generative models, neural network), non-parametric approaches (k-Nearest Neighbor), and cross-validation.

7. Course logistics:

There will be weekly homework (40%), a take-home midterm (30%), and a final project (30%). Both the take-home midterm and the final project will include questions asking you to implement some algorithms and perform experiments using either Python or MATLAB (you can choose). The final project will be a guided project that will provide you with an opportunity to develop and evaluate statistical inference techniques for real-life applications.