This in-class tutorial/homework will use MATLAB to understand ROC analyses more deeply.

**1) Explore the relationship between signal and noise distributions and the ideal observer’s ability to detect the signal. (75 pts)**

Open the **ROC\_hw1\_script.m** file from the Canvas site in MATLAB. You will complete the STEPS and QUESTIONS below for each of the following three different types of distributions. Please show your work by pasting figures and providing annotation of what you see/did.

1. **Normal (25 pts)**
   1. For noise mean=12 & std=2, signal mean=18 & std=2, the ROC area is 0.9119.
   2. For noise mean=11 & std=2, signal mean=13 & std=2, the ROC area is 0.7602.

For noise mean=11 & std=2, signal mean=12.9 & std=2, the ROC area is 0.7491.

For normal distribution, delta mean = 1.9 for ROC area to be around 0.75.

* 1. Normal distribution is usually applicate to sensor noise for signal and image processing, where the background noise is often modeled to be gaussian noise. A detection task uses this distribution to model could involve distinguishing a weak target signal (signal mean & noise mean difference is not significant) embedded in Gaussian noise.

1. **Poisson (25 pts)**
   1. For noise mean=21 and signal mean=60, the ROC area is 0.9979.
   2. For noise mean=21 and signal mean=42, the ROC area is 0.9897.

For noise mean=21 and signal mean=25.7 the ROC area is 0.7531.

For normal distribution, delta mean = 4.7 for ROC area to be around 0.75.

* 1. Poisson distribution is usually applicate to neuronal spike train modeling in neuroscience, where the event happens discretely over a fixed interval. A detection task uses this distribution to model could involve detecting the presence of a stimulus based on spikes recorded from neurons follows a Poisson distribution.

1. **Rayleigh (25 pts)**
   1. For noise mean=5 and signal mean=10, the ROC area is 0.7997.
   2. For noise mean=5 and signal mean=20, the ROC area is 0.9313.

For noise mean=5 and signal mean=8.6 the ROC area is 0.7473.

For normal distribution, delta mean = 3.6 for ROC area to be around 0.75.

* 1. Rayleigh distribution is usually applicate to wind or wave processing, where the event is continuous with independent normal variable and equal variance. A detection task uses this distribution to model could involve detecting the presence of a cluttered/scattered stimulus raised from multiple independent sources combined.

**STEPS:**

1. Read through the script. Note that there are 3 sections at the top to define the probability distributions for noise and signal. Use the relevant section by uncommenting that section (Ctrl-T on PC), making sure the other sections are commented out (Ctrl-R). Note that you can play around with the resolution (event\_resol) and the min/max of the event alphabet to best suit your distribution (i.e. the distribution fills most of your event range).
2. Set a mean for the noise and the signal distributions (change from the default). You will rerun all steps for 3 different sets of values for each distribution.
   1. As a starting point, choose reasonably well separated means so that you can see that the noise and signal distributions overlap to some degree, but are not completely on top of each other, and are not completely separate
3. Run the entire script. This will generate the distributions, plot the log-likelihood ratios, from which an Ideal Observer can make decisions with different threshold criteria, and then plot the resulting Receiver Operating Characteristics curve.
   1. the first shows the probability densities for the assigned distribution (black = noise; red = signal)
   2. the second shows the log-likelihood as a function of the event, with coloring showing which events have been assigned to the noise (black), and which are assigned to the signal (red), based on a likelihood criterion, k (see “Compute Ideal Observer-based ROC curve” section of script).
   3. the third shows the ROC curve plotted for different values of the likelihood criterion, k. You can adjust the spacing of the “x’s” plotted by changing the resolution and limits for k in the for loop (k\_values variable).

**For each of the two distributions, answer the following:**

**QUESTION (10 pts):** For a particular choice of noise and signal mean, measure the Area Under the ROC Curve (AUC). Use the trapz() function in MATLAB if it is available in your version. If not, use the “data cursor” in the Tools of the figure to read ~10 points from which you can estimate the area yourself.

**QUESTION (10 pts):** Repeat for two other combinations of means. For the distribution type, how different do the means need to be in order for the area under the ROC curve to be ~0.75?

**QUESTION (5 pts):** For the distribution you are using, research what types of applications they are usually found in. What might be a type of detection task that would give rise to this kind of noise and signal distributions?

**2) ROC analysis to characterize neuron’s detection of sound based on spike counts. (25 pts)**

In the plots below for two different neurons, you can see the raster (individual dots over 180 trials) and PSTH (histograms) plots of spiking activity over a 600 ms window, with sounds presented from 200 to 260 ms. The interval from 0 to 200 ms can be considered the spontaneous activity of the neurons.

**Unit 3065: spike counts measured from 200-300 ms**



**Unit 3009: spike counts measured from 200-400 ms**

To start, you should download the **ROC\_hw2\_script.m** file from the Canvas site, and navigate to the resulting folder within MATLAB. Also download the **ROC\_hw\_data.mat** file, which contains the spike counts during spontaneous and sound-evoked periods (see figures above) for the two units.

Read through the script – note the similarity to the in-class exercise. Note that there are 2 sections near the top to choose which units you will analyze when you run the script. Switch between units by uncommenting that section (Ctrl-T on PC), making sure the other section is commented out (Ctrl-R).

**QUESTION (5 pts):** Looking at the firing rates and patterns of the two different units, do you expect the units’ ability to detect the sound compared to silence (i.e. spontaneous activity) based on spike counts is likely to be similar or better for one unit or the other. Provide a justification.

Looking at firing rate of two units, unit 3065 fires more rapid and transient with lower baseline firing rate, while unit 3009 fires slower and more sustained with higher baseline firing rate. I expect unit 3065 is better at detect sound compared to unit 3009 because its lower baseline firing rate allows for higher “change” (or z-score firing rate / signal-to-noise ratio) when responding to a sound stimulus, thus make it easier to differentiate from silence and have higher sensitivity to signal.

**QUESTION (20 pts):** Run the script to plot the distribution of spike counts, the likelihood ratios, and the ROC curves for the two different units. Explain what you see, comparing the distributions of spontaneous and sound-driven spike counts. Compute the area under the ROC curves for each and compare. How does this compare to your expectation?

A graph with numbers and symbols

AI-generated content may be incorrect.A graph of events and events

AI-generated content may be incorrect.Unit 3065 (ROC area = 0.7723)

Unit 3009 (ROC area = 0.8443)

A graph of a number of data

AI-generated content may be incorrect.A graph of events and events

AI-generated content may be incorrect.

For unit 3065, the neuron have highest spike probability at event 0, suggests its most frequent response occurs at the onset of the stimulus; whereas unit 3009’s highest spike probability is around event 2, meaning its strongest response occurs slightly later after the onset of stimuli. This observation is expected because unit 3065’s response is rapid and transient whereas unit 3009 is slow and sustained.

Looking at likelihood ratio, both neurons have their highest log likelihood around event 4, but unit 3065 reach the value of 3, while unit 3009 reach 2, this suggests that unit 3065’s response to stimuli is more distinguishable from its baseline (more sensitive). This is expected because unit 3065 have slightly lower baseline firing rate that would allow for greater distinguishability between stimuli and noise.

Finally, by evaluating ROC area (area under curve), we observed that unit 3065 have smaller area than unit 3009, which suggests that unit 3009 can better distinguish between sound and silence. This is surprising because we expected unit 3065 to be better at detecting sound because of its lower baseline firing rate and sharp response pattern.

This may be due to that for rapid-response unit like 3065, its response is highly concentrated within the first 20-30ms of the 60ms analysis window, making its signal more localized in time. In contrast, for sustained-response unit like 3009, the response is consistant throught the window, which may lead to better overall separability between stimulus driven and spontaneous activity. This may also suggests that while rapid, transient responses can be highly sensitive to stimulus onset, a more prolonged response pattern may provide greater reliability in distinguishing sound from silence over a longer time window (greater ROC area).