

Assignment A2a: Signal Detection

Put your Name and Case ID here

Use `464-A2a-yourcaseid.ipynb` for your notebook name.

If you will use your own data files, put these in a folder named `464-A2a-yourcaseid-files` , which should also contain your notebook. Make sure this folder contains all the files needed to reproduce your results.

Overview

This assignment focuses on detecting simple signals in noise.

Readings

The following material provides both background and additional context. It linked in the Canvas page for this assignment.

- Dusenbery, D. B. (1992). *Sensory Ecology*. Chapter 5 Signal Detection, sections 5-1 and 5-2.

Learning objectives

- write code to generate random signals
- use vector operations and logical indexing to concisely express computational ideas
- measure different types of detection errors
- characterize different types of error profiles with ROC curves

Exercises

1. Generating signals with events and additive noise

1a. Signals in Gaussian noise

Write a function `genwaveform(N=100, α=0.1, A=1, σ=1)` to generate a waveform that is a linear combination of a sparsely occuring signal and additive Gaussian noise. The parameters specify the waveform length `N` , the signal probability `α` , the signal amplitude `A` , and noise standard deviation `σ` . Assume the noise mean is zero. The values listed are defaults. The signal probability specifies the probability of an event occurring within a sample. Assume the events are independent. The function should return a tuple of the resulting waveform and array of the event locations as indices.

Plot the generated signal and display the location of the events with markers.

(Comment on terminology: The term "signal" can refer either to an individual event or the collection of events as a whole. The waveform is the signal plus the noise. Note that "signal" is sometimes used loosely to refer to the observed waveform, rather than the waveform without the noise. This is because the signal itself cannot be observed directly, only inferred. The term "underlying signal" is often used to emphasize the component of the waveform without the noise.)

1b. Signals in uniform noise

Modify the `genwaveform` function so that it accepts an argument `noisetype` to specify the type of noise. Here we will use `Gaussian` and `uniform` . For uniform noise, we again assume zero mean. The `σ` parameter should be interpreted as the width of the uniform distribution with range $[-σ/2, σ/2]$.

Plot an example using uniform noise.

1c. Notation

Mathematically, the expression for the observed waveform $y(t)$ is written as the sum of the signal $x(t)$ and the noise $ε(t)$: $y(t) = x(t) + ε(t)$.

Write two equations to express 1) the signal in terms of N events that occur at times t_i and 2) the noise as being distributed according to a Normal with mean $μ$ and variance $σ^2$.

1d. Conditional probability

What is the expression for the probability of the waveform at time t given that there is a signal?

2. Signal detection

2a. Effect of parameters on detection probability

Explain what effect the parameters and type of noise have on detection probability. For what values does the probability reduce to pure chance? Or become certain (i.e. approach 1)? Explain your reasoning and illustrate with plots.

2b. Types of detections and detection errors

Write a function `detectioncounts(si, y, θ)` which given an array `y` , signal index `si` , and threshold `θ` , returns a named tuple (tp, fn, fp, tn) of the counts of the true positives, false negatives, false positives, and true negatives.

Write a function that plots the samples and threshold and shows the true positives, false negatives, and false positives with different markers.

2c. Detection probabilities

What is the mathematical expression for the probability a false positive? What is it for a false negative? (Note that these are conditioned on the signal being absent or present, respectively.)

Write the functions `falsepos` and `falseneg` to return the expected false positive and negative rates. The first argument should be the threshold $θ$, the rest of the arguments should be keyword arguments that follow those of `gensignal` but without unnecessary parameters.

What are the expected error probabilities using the information and count results from above? How could you estimate these from the distribution parameters and detection threshold? Show that your empirical results consistent with those calculated analytically.

3. ROC cures

3a. Threshold considerations

Explain why, in general, there is not an optional value for the threshold. What value minimizes the total error probability? How is that different from minimizing the total number of errors?

3b. ROC plot

Write a function `plotROC` to plot the ROC curve using the functions above. It should use a similar parameter convention.

Tests and self checks

You should write tests for your code and make plots to verify that your implementations are correct. After you submit your draft version, take the self check quiz. This will give you feedback so you can make corrections and revisions before you submit your final version. Here are examples of the types of questions you can expect

- conceptual questions from the readings and lectures
- questions from the assignment
- plot waveforms of signals in Gaussian and uniform noise using specified parameters
- plot examples that have high and low SNR
- question that use reference data ("A2a-testdata.h5" in "Files/assignment files" on Canvas)

Submission Instructions (up to a -5 pt penalty)

- Restart your kernel and re-run your notebook.

Verify that each cell produces the correct output. The cells should be numbered sequentially starting from `[1]` . This will help ensure that your notebook always produces the output you intended and will help avoid errors arising from non-sequential cell execution and inconsistencies in variable definitions.

- Export your notebook to pdf.

This provides a static rendering of your notebook and improves the grading workflow. Currently, a pdf file is the only way to view your notebook submission within canvas. Refer to "Notebook export tips" on Canvas/Files for how to get good pdf output.

- Submit the following files to canvas:

- `464-A2a-yourcaseid.pdf`
- `464-A2a-yourcaseid.ipynb`

If your notebook relies on your own data files, in addition to the files above, also submit a zip file named `464-A2a-yourcaseid-files.zip` of a folder containing the notebook and datafiles. Your zip file should unzip to a folder named `464-A2a-yourcaseid-files` . Do not use other compression formats. Do not submit large (> 10 MB) data files. Instead, provide a link where they can be downloaded.