Quantitative Methods in Neuroscience
(NEU 466M)
Homework 3
Due: Tuesday Feb 18 by 12 pm (uploaded to Canvas)

In this assignment you will explore spike train statistics. General guidelines: Read through each complete problem carefully before attempting any parts. Feel free to collaborate in groups of size 2-3, but always note the names of your collaborators on your submitted homework. For graphs: clearly label your axes and use good color and symbol choices. Print out your matlab code (in the form of a script file). For derivations you're asked to do 'by hand' (in other words, analytically, using paper and pencil) feel free to turn in handwritten or typed-out work.

1) **Loading/reading datafiles in Matlab; a first characterization of the blowfly H1 neuron for horizontal motion estimation.**

   a. Download the data file `c1p8.mat` from the course webpage and save to a local directory. Load this data into matlab using the command `load` on the command line or by navigating to the 'Open' menu option on your Matlab work window. This data file has spikes from the H1 neuron (`rho`) in response to the stimulus (`stim`), both sampled at 500 Hz for a total of 20 minutes. Use `whos` to check the variable names and dimensions. Compute the mean of `stim`, `rho`. Compute the standard deviation of `stim`. Is it fair to say the mean is 'small' or 'close' to 0 or not? How would one set an overall scale for comparison, to determine what mean value counts as 'small'? Compute and plot a histogram of `stim`, with 40 bins (use `hist`).

   b. Compute the autocorrelation function of `rho` using `xcorr` and plot it (zoom using `xlim`). Mark the features of the autocorrelation, and write on the plot, in real units (seconds or milliseconds), the widths of each of the features: 1) the central peak, 2) the narrow dip around the peak, and 3) the slower hill around the central peak and dip. Do the same for `stim`.

   c. What relevant quantity does the height of the central peak of the autocorrelation of the spike train represent? Derive/prove this claim by hand (pencil and paper), using the autocorrelation formula and any relevant properties of the spike train.

   d. Is it ever possible for (any parts of) the autocorrelation function of a spike train to become negative? Prove your answer by hand. Is it ever possible for a zero-mean signal to have negative autocorrelation values? Show how this could be by hand – i.e., construct a simple example of a signal to support your answer. Matlab:

What changes if you first mean-subtract the spike-train `rho` before computing its autocorrelation?

e. Plot the cross-correlation function $C_{\text{stim, rho}}$. Explain why the peak is where it is. Interpret this plot in detail in terms of what it says about changes in the stimulus and the spike response. For example, "whenever the stimulus becomes positive, the response $\cdots$ "'. Now plot $C_{\text{rho, stim}}$. Interpret and explain in similar terms: "whenever the cell spikes, $\cdots$".[1]

2) **Analysis of structure in spike train data from the entorhinal cortex: autocorrelation, crosscorrelation, and interspike histograms** Entorhinal cortex exhibits prominent oscillations in the collective activation of its neurons. These oscillations include prominent contributions in the theta (6-12 Hz) and gamma (40-100 Hz) bands. However, it is difficult to see these oscillations by eye in the spike trains of single neurons. Our aim is to mine neural spike train data for signatures of oscillations. As in class, we'll look at crosscorrelations, and then we'll also look at interspike interval histograms to see if certain interspike intervals are more frequent than others.

a. Load `gridcell_halfmsbins.mat`, a data file with spike trains of three simultaneously recorded entorhinal cortical cells from the same electrode bundle (thus, these cells are very close to each other, within about a 100μm radius [2]). 1's represent a spike in the given time-bin, and the data are sampled at 2000 Hz. Call these trains `s1`, `s2`, and `s3`, respectively, and call the length of each train N. Scrutinize by eye the spike train `s1` in the appropriate time-windows to look for signatures of theta- or gamma-band oscillations, make a plot, and explain your result. Compute the autocorrelation of `s1`, and plot it around with 1000 time-bins on each side of the central peak (you can compute the full autocorrelation then use `xlim`, or restrict `xcorr` using `maxlag`). Identify the refractory period and anything else that appears interesting. Zoom in or out to search for oscillations, and print out plots with relevant features. Note the duration of the features in milliseconds.

b. Assessing the significance of structure in the autocorrelation: How do we know when a peak or a dip in an autocorrelation is significant or due to noise? A chance noise-driven feature is more likely when the signal is short: why? One way to

[1] We will revisit the quantity $C_{\text{rho, stim}}$ in future lectures because it has a very interesting and useful interpretation for trying to solve the *encoding* problem.

[2] If you're curious about the estimated distance in cortex between cells recorded on an electrode, see Mechler et al., Three-dimensional localization of neurons in cortical tetrode recordings, J. Neurophysiol., 2011.

assess the significance of features in the autocorrelation is to generate random spike trains with the same number of spikes as the original train, and compare its autocorrelation with the original one. Randomize `s1` as follows: create a randomly ordered vector of indices from 1 to N by `randind=randperm(N)`. Then `s1rand= s1(randind)` is a randomly permuted (ordered) version of the spike train. Plot autocorrelations of `s1` and `s1rand` on the same plot. Which features of `s1` are not present in `s1rand` and which features continue to be present? The random spike train autocorrelation gives you a measure of the level of autocorrelations you would expect in a spike train with no specific structure – i.e., by chance. Values above it can be deemed significant based on how much higher they are from this assessment of chance.

c. Shuffling spike trains: The previous randomization method preserved only the total number of spikes over the recorded interval (the mean spike rate), while removing (randomizing) all structure in the spike train. The next step is to ask whether there is important structure in the data beyond adjacent spike pairs – thus, we will generate a partially randomized dataset that preserves inter-spike intervals of adjacent spikes, but randomizes the ordering of the intervals. (The result will be a spike train in which every interspike interval will occur with the same frequency as in the data-set, but the ordering of events will be randomized.) Here's how: Compute the list of spike times in `s1` using `find`. Use `diff` on the list of spike times to generate a list of interspike intervals. Randomly permute the list of interspike intervals. Finally, reconstitute a list of spike time indices from the permuted intervals (hint: use `cumsum`). Finally, generate a vector of zeros of the same size as $s1$, and set all entries at the list of reconstituted spike time indices to 1. You now have a spike train `s1shuff` that preserves the mean spike count and the interspike intervals in `s1` (we'll verify this in e. below), but all other structure is randomized.

   Compare the autocorrelation of `s1` and `s1shuff` on the same plot. What has changed and what has remained the same? Think (hard!) about this and explain why.

d. Now compute the cross-correlation between all pairs of `s1, s2, s3` and make separate plots for each. Superimpose the cross-correlations between the corresponding pair of `s1rand, s2rand, s3rand` and of `s1shuff, s2shuff, s3shuff` on the corresponding plot. Compare and contrast with the autocorrelation results and for each feature, and explain.

e. Interspike interval histogram. Another way to look for temporal structure in neural spike data is to analyze the histogram of interspike intervals. Plot his-

tograms of the interspike intervals for each of `s1, s1rand, s1shuff` (use `hist` with 1000 bins). Zoom to small intervals, and identify peaks/plateaus that might correspond to theta- and gamma-band oscillations.