MA 568 – Statistical Analysis of Point Process Data

Problem Set #1
Due October 2, 2018

Cells in the brain communicate and perform computations using electrical impulses, called spikes. Point process models are often used to propose an explicit relationship between spikes recorded from the brain and other measured signals including sensory stimuli and motor behavior. For many neuroscience experiments, an important first task is to characterize the 'baseline' spiking activity that is observed in the absence of interesting stimuli. A common assumption about this baseline activity is that it can be described as a simple Poisson process.

In this problem, we will attempt to characterize the firing properties of the maintained discharges of retinal ganglion cells of the goldfish under constant light and environmental conditions. This data is described in more detail in Iyengar & Liao, *Biological Cybernetics*, 1997.

1. Please download the **`Retinal_ISIs.mat`** (or **`Retinal_ISIs.txt`**) file from the course website. This file contains the waiting times, or interspike intervals (ISIs), in milliseconds, of a single retinal ganglion cell over 30 seconds. Open this file and plot the spiking activity as a histogram of the distribution of the times between spiking events and as a spike train time series. Describe the spiking properties of this neuron based on your visualizations.

Useful MATLAB and (R) functions: **`hist, cumsum`**

2. Compute a 5-number summary (minimum, .25 quantile, median, .75 quantile, and maximum) and a box plot for the interspike interval (ISI) distribution. What do these statistics tell you about the structure of the data?

Useful MATLAB and (R) functions: **`sort, boxplot`**

3. Another common approach to analyzing spiking data is to discretize time into bins of fixed width and count the number of events that occur in each time bin. Bin the spike train data from **`Retinal_ISIs.mat`** into time bins of width 1 ms., 10 ms., and 100 ms.. Plot both the time series of spike counts and the distribution of spike counts as a histogram for each bin width. How do these distributions compare to the model for the increments of a Poisson process?

4. In class, we showed that the log likelihood for a homogeneous Poisson process with rate parameter λ is given by:

$$\log L(\lambda) = N(T)\log(\lambda) - \lambda T,$$

Where N(T) is the number of spikes in the observation interval [0,T]. Plot the likelihood as a function of λ for values of λ between 0 Hz to 60 Hz. Find the value of $\hat{\lambda}_{ML}$ that maximizes the likelihood. Provide an approximate 95% confidence interval for $\hat{\lambda}_{ML}$.

Goodness-of-fit:

In class, we discussed a number of approaches for determining the goodness-of-fit of point process data to a Poisson model. These include tests of the distributions and of the independence of the interspike intervals and increments process.

5. Plot an empirical CDF of the interspike intervals for the data. For a Poisson process, the theoretical interspike intervals should be independent exponentially distributed with parameter $\hat{\lambda}_{ML}$. Plot this exponential CDF on the same plot as your empirical CDF. Compute the KS statistic, defined as $\max|F_{Emp}(x) - F_{model}(x)|$. Construct a KS plot of the empirical CDF on the x-axis against the model CDF on the y-axis, including 95% confidence bounds, given approximately by $\pm 1.36/\sqrt{n}$, where n is the number of observed ISIs.

6. Construct a QQ plot of the empirical vs. model quantiles. For which ISI values does the QQ plot most deviate from expected?

Useful MATLAB and (R) functions: **`expinv, (qexp)`**

7. Compute the sample Fano Factor for the increments process binned at 1 ms, 10 ms, and 100 ms resolutions. The sample Fano Factor for a Poisson process is distributed as a $\Gamma\left(\frac{n}{2}, \frac{2}{n}\right)$ random variable. Do these statistics fall into the 95% confidence interval for a Poisson process at each of these bin lengths?

Useful MATLAB and (R) functions: **`gaminv, (qgamma)`**

8. Plot the autocorrelation function of the observed interspike intervals with 95% confidence bounds. The 95% confidence interval for the correlation coefficient of independent Gaussian random variables is $\pm 1.96/\sqrt{n}$. What does this suggest about the assumption that the ISIs are independent?

Useful MATLAB and (R) functions: **`expcdf, xcorr, (pexp, acf)`**

9. What conclusions can you draw about the spontaneous firing properties of these retinal ganglion neurons? Given that environmental conditions for this experiment are constant, what other variables could be added to our baseline neural model to improve goodness-of-fit?