Chapter 2: Introduction to Point Processes

I. Point processes are used to describe data that are localized in space or time

In Chapter 1, we saw an example of neuronal activity in the supplemental eye field (SEF) expressed in terms of a raster plot and a peri-stimulus time histogram (Fig. 1.1). The raster plot shows locations of action potentials in time for multiple trials, and the peristimulus time histogram counts the number of such events is small time bins, averaged over all of the trials. These types of plots provide a means to express data that consists of discrete events localized in time.

Analyzing data of this sort presents its own unique challenges, and poses its own set of questions. What are the different ways to describe the data? What types of stochastic models are appropriate for explaining the structure in the data? How can we measure how well the data is described by a particular model? In order to address these questions, we require a specific mathematical structure that can handle data of this sort.

A temporal **point process** is a stochastic, or random, process composed of a time-series of binary events that occur in continuous time (Daley and Vere-Jones, 2003). They are used to describe data that are localized at a finite set of time points. As opposed to continuous-valued processes, which can take on any of countless values at each point in time, a point process can take on only one of two possible values, indicating whether or not an event occurs at that time. In a sense, this makes the probability models used to describe point process data relatively easy to express mathematically. However, point process data are often inappropriately analyzed, because most standard signal-processing techniques are designed primarily for continuous-valued data. A fundamental understanding of the probability theory of point processes is vital for the proper analysis of point process data.

The study of point processes is especially crucial for neural data analysis. Brain areas receive, process, and transmit information about the outside world via stereotyped electrical events, called action potentials or spikes. Spikes are the starting point for virtually all of the processing performed by the brain.

The firing properties of many classes of neurons are known to correlate to specific extrinsic signals such as sensory stimuli or behavioral or motor outputs. For example, auditory signals are transduced by the cochlea into spiking patterns in collections of neurons, each of which respond to a particular frequency. These correlations are present not only in sensory neurons near the periphery of the CNS but also in those that are many synapses removed from the periphery. For example, cells in the CA1 region of the rat hippocampus exhibit place field structure, whereby their firing properties correlate with the animal's location within its environment. Similarly, in motor cortex, the plan for a complex body movement is represented in the ensemble firing of neurons tuned to various kinematic and kinetic features of the desired movement. The fact that neurons over a broad range of neural systems represent information about external biological and behavioral signals gives rise to the concept of neural coding. Under this viewpoint, neural firing is viewed as a type of coded language from which outside observers of the spike train sequences could decode information about the outside world, if only they had appropriate neural models with which to decipher these signals.

Therefore, cracking the neural code involves studying the relation between brain signals and these external biological signals.

Figure 1. Example of spiking activity of a neuron in the human subthalamic nucleus. (This is a placeholder figure only)

The timing of spiking activity is related to an underlying membrane voltage process that is typically not recorded for *in-vivo* experiments. Figure 1 is an example of a typical extracellular voltage trace showing the spiking activity of a single neuron. The stereotyped nature of these action potentials suggests that the information contained in sequences of spiking activity, or spike trains, is not related to the shape of the voltage trace for any particular spike, but rather to the frequency and timing of these events. At the same time, a neuron's responses to repeated presentations of the same stimulus are stochastic. That is, with multiple presentations of a stimulus to a neuron or ensemble, the set of resulting spikes will differ in their exact timing, although they may share common statistical features. In some brain systems, information about an encoded signal can be transmitted in a small number of spikes or in the exact arrival times of these events. Taken together, these properties of neural spiking suggest that they are most appropriately modeled as point processes.

Example 1. Retinal Neuron Under Constant Light and Environmental Conditions

Neurons in the retina typically respond to patterns of light displayed over small sections of the visual field. However, when retinal neurons are grown in culture and held under constant light and environmental conditions, they will still spontaneously fire action potentials. In a fully functioning retina, this spontaneous activity is sometimes described as background firing activity, which is modulated as a function of visual stimuli.

Figure 2 shows the spiking activity of one such neuron firing spontaneously over a period of 30 seconds. Even though this neuron is not responding to any explicit stimuli, we can still see structure in its firing activity. Although most of the interspike intervals are shorter than 20 msec, a significant fraction of these ISIs are much longer, on the order of 60-120 msec. We can also observe bursts of firing with multiple spikes arriving in quick succession of one another.

Spontaneous spiking activity that does not clearly relate to any external biological or behavioral signals is useful for constructing simple models for how each spike relates to the neuron's past spiking history, and can help us understand the fundamental biophysical properties of action potential generation. We shall see that history dependence is an important component of virtually all neural spiking activity and that accurate models of history dependence are essential in fully describing most spiking data.

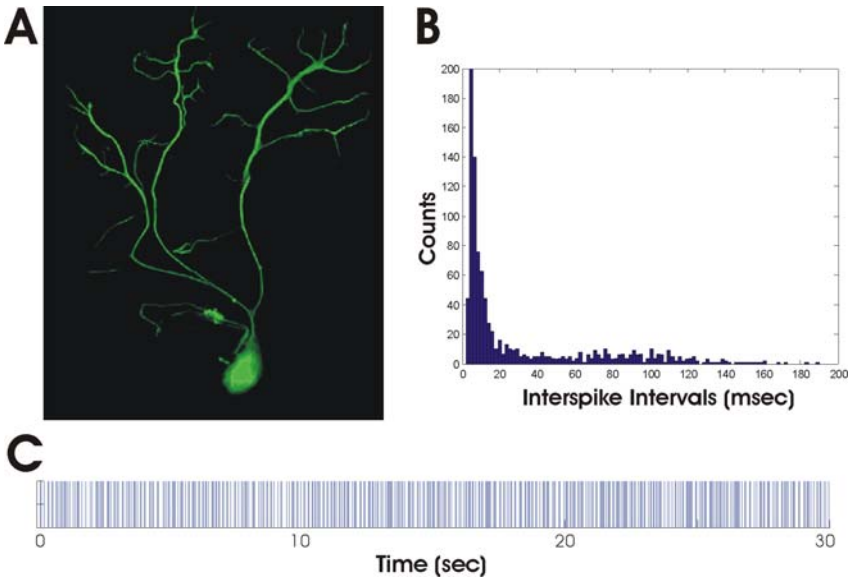

Figure 2. Spontaneous spiking activity of a goldfish retinal neuron in culture under constant light and environmental conditions over 30 seconds. A) Retinal ganglion cell (taken from web, may be copyrighted) B) Histogram of interspike intervals. C) Spike train - times series of spiking data.

Example 2. Spiking activity of a primary motor cortical (M1) Neuron

The spiking activity of neurons in primate motor cortex has been shown to relate to intended motor outputs, such as limb reaching movements. Experiments where a monkey performs a two-dimensional reach have shown velocity dependent cosine tuning, whereby a motor cortical neuron fires most when the hand moves in a single preferred direction and the intensity drops off as a cosine function of the difference between the intended movement and that preferred direction, and additionally increases with increasing movement speed. Figure 3 shows an example of the spiking activity of a neuron in primate motor cortex as a function of hand movement direction during a center-out reaching task. The neuron fires most intensely when the hand moves in a direction about 170 degrees from east. These firing patterns have also been shown to vary as a function of movement speed (Moran & Schwartz, (1999))

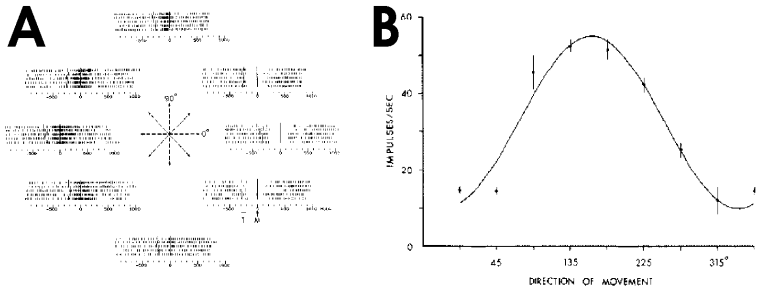

Figure 3. Cosine tuning in primate motor cortex. A) Spike rasters for a center out task with eight principal directions. B) Spike count as a function of direction shows a sinusoidal trend. (Placeholder figure, may be copyrighted)

Example 3. Conditional Intensity Function for Hippocampal Neuron

Neurons in rodent hippocampus have spatially specific firing properties, whereby the spiking intensity is highest when the animal is at a specific location in an environment, and falls off as the animal moves further away from that point. Such receptive fields are called **place fields**, and neurons that have such firing properties are called **place cells**. Figure 4 shows an example of the spiking activity of one such place cell, as a rat executes a free-foraging task in a circular environment. The rat's path through this environment is shown in blue, and the location of the animal at spike times is overlain in red. It is clear that the firing intensity is highest slightly to the southwest of the center of the environment, and decreases when the rat moves away from this point.

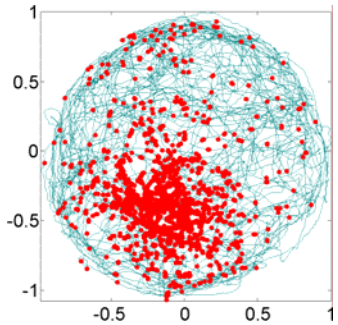

Figure 4. Movement trajectory (blue) and hippocampal spiking activity (red) of a rat during a free-foraging task in a circular environment.

Neural spike trains are described by temporal point processes because the spike events are localized in time. It is also possible to use point process theory to model data that is localized at a discrete set of locations in space or in both space and time. These models are called spatial and spatiotemporal point processes respectively. The spike sorting problem, that is, the problem of determining how many neurons are contributing to the firing activity recorded on a set of electrodes and assigning each spike to the neuron that generated it, can be considered a clustering problem on a spatial point process. The space in which these spikes occur is a high-dimensional abstract space describing possible spike waveforms. Each individual spike has a particular waveform, and the overall spiking activity from a single cell will cluster in a small region of this space. By modeling the probability of spiking for each neuron, electrophysiologists can assign each spike to a single neuron and determine their confidence in the classification results.

Here, we focus on theory related to temporal point processes. Unless otherwise noted, when we refer to a point process it will be assumed to relate to spiking activity in time. The theory and methods associated with spatial and spatiotemporal point processes are analogous to those of pure temporal point processes. When indicated, we will point out extensions of the theory as they apply specifically to spatial point processes.

Another neuroscience application that makes good use of point process methods is nuclear medicine imaging, such as positron emission tomography (PET). To perform a PET scan, clinicians introduce a radioisotope that has been incorporated into a metabolically active molecule into the patient's bloodstream. These molecules become concentrated in specific tissues and the radioisotopes decay, emitting positrons. These

emissions represent a spatiotemporal point process since they are localized occurrences throughout the tissue and in time. After being emitted, these positrons interact with nearby electrons, producing a pair of photons that shoot out in opposite directions and are detected by a circular ring of photosensors. The arrival of photons at each sensor represents a temporal point process, and by characterizing the temporal interactions between arrivals at multiple sensors, it is possible to infer the original location of the positron emission. By observing and inferring the locations of many such occurrences, it is possible to construct an image of specific metabolically active tissues.

The theory of point processes has also been applied to many physical phenomena outside of the neurosciences. For example, temporal point processes have been used to characterize the timing and regularity of heart beats (Barbieri and Brown, 2005). In geology, point processes have been used to describe geyser eruptions (Azzalini and Bowman, (1990)). Spatiotemporal point processes have been used to characterize and predict the locations and times of major earthquakes (Ogata, 1988). For each of these processes as is true for neuronal spike events, there is an underlying continuous-valued process that is evolving in time and the associated point process event occurs when the underlying continuous process crosses a threshold. The continuous process for the electrical event preceding the contraction of the ventricles is like for a neuronal spike event, a subthreshold membrane voltage. The continuous process underlying a geyser eruption is water pressure, whereas the continuous process underlying an earthquake is pressure between the tectonic plates on either side of a geological fault line.

II. A point process may be specified in terms of spike times, inter-spike intervals, or spike counts.

As with any random signal, we can express the data we observe equivalently in terms of multiple collections of random variables. Let's take a closer look at some of the different ways of describing point process data. As is standard with probability theory, we will use capital letters to indicate random variables and lower case variables to indicate the data values that they take, unless otherwise indicated.

Let $S_1, S_2, \ldots$ be random variables describing the occurrence times or spike times of a point process. A realization of a point process is the event $S_1 = s_1, S_2 = s_2, \ldots$ for some collection of times $0 < s_1 < s_2 < \ldots$. Let $X_1, X_2, \ldots$ be a set of random variables describing the possible waiting times between occurrences, or inter-spike intervals. We can compute the waiting times from the spike times by taking the difference between subsequent spike times. Mathematically, $X_1 = S_1$ and $X_i = S_i - S_{i-1}$. Similarly, we can compute the spike times by taking the cumulative sum of all the waiting times. That is, $S_n = \sum_{i=1}^{n} X_i$. Clearly, there is a one-to-one correspondence between any set of spike times and any set of inter-spike intervals.

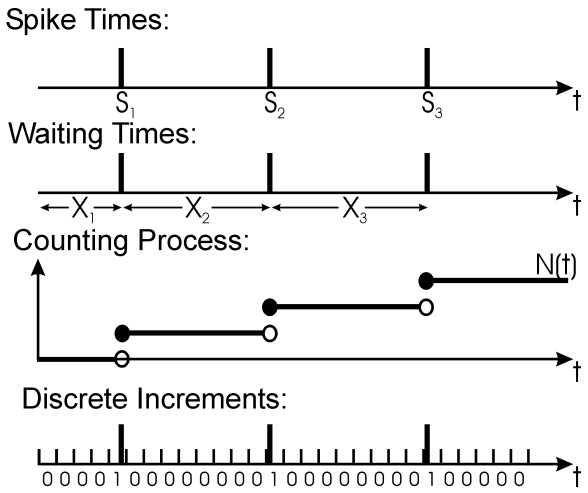

Figure 5. Multiple specifications for point process data.

Another useful way to describe a set of spiking observations is in terms of the number of spikes observed over any time interval. We define $N(t)$, the **counting process**, as the total number of spikes that have occurred up to and including time t (Fig. 5). The counting process gives the number of spikes observed in the interval $(0,t]$. If we let $\Delta N_{(t_1,t_2)}$ denote the total number of spikes observed in an arbitrary interval $(t_1,t_2]$, then we can compute this from the counting process, $\Delta N_{(t_1,t_2)} = N(t_2) - N(t_1)$. $\Delta N_{(t_1,t_2)}$ is sometimes called the **increment** of the point process between t_1 and t_2. We see that keeping track of the times at which the counting process increases, is equivalent to keeping track of the spike events. Therefore, characterizing the spike events is equivalent to characterizing the counting process, and vise versa.

For both of the spike time and the inter-spike interval process, the index of the variable is discrete, indicating to which spike event we are referring, while the set of possible values that the variable can take on is continuous. S_1 refers to the time of the first spike, and can take on any value in the range $[0,\infty)$. The counting process, $N(t)$, on the other hand, has a continuous index, indicating a point in time, while the set of possible values that it can take at any instant are discrete integers. Clearly, then, the probability distributions associated with these different random variables will take different forms. However, we can use these different variables to express the same events. For example, the set of times $\{t : N(t) < j\}$ when the counting process is less than some value j is equivalent to the set of times $\{t : S_j > t\}$ when the j^{th} spike has not yet occurred. Both of these express the set of all times that precede the j^{th} spike, but they do so using distinct random variables. Specifying any one of the spike times, inter-spike intervals, or counting process fully specifies the other two and fully specifies the point process as a whole. It is useful to have multiple equivalent expressions for the same random process so that if the probability distribution used to describe one data set is complicated, we can transform the data set into something easier to work with, without losing any information about the process.

The spike time, inter-spike interval, and counting processes are all continuous time specifications of a point process. It is often useful, both for developing intuition and for constructing data analysis methods, to consider point processes in discrete time. One approach for constructing a discrete-time representation of a point process is to take the observation interval $(0,T]$ and break it up into n small, evenly spaced bins. Let $\Delta t = T/n$, and $t_k = k \cdot \Delta t$, for $k = 1,...,n$. We can now express a spike train in terms of discrete increments $\Delta N_k = N(t_k) - N(t_{k-1})$, which count the number of spikes in a single bin. If Δt is small enough so that the process cannot fire more than one spike in a single bin, then the set of increments $\{\Delta N_k\}_{k=1,...,n}$ is just a sequence of zeros and ones indicating in which bins spikes are observed (Fig. 5).

Whereas the continuous time specifications for a spike train contain all of the possible information about the process, this discrete time description loses information about the precise timing of spikes within a bin. However, in the limit as the discrete-time bin size, Δt, goes to zero, the discrete-time increments become as informative as the continuous-time descriptions, and the likelihoods, probability distributions, and estimation algorithms associated with the discrete-time methods converge to their continuous-time counterparts. In many cases, the theory and analysis methods will be easier to develop and implement in discrete-time.

III. Point processes can display a wide variety of history-dependent behaviors.

We will characterize different classes of point processes based on the probability of firing a spike in a small time bin. One important feature of any particular point process is how the history of past firing activity affects the probability of firing a spike now. In some cases, especially when the occurrences are rare compared to the time scale of the underlying process or the effect of history has been averaged out by combining multiple point processes, Poisson processes can accurately describe spiking activity. For example, when the spontaneous spiking activity from multiple neurons is recorded on a single electrode and left unsorted, the past firing of one neuron may have no effect on the firing probabilities of the other neurons, so that the combined spiking activity shows little to no history dependence.

However, in many physical systems that produce point process data, history dependence in an expected consequence of the mechanisms underlying the generation of each occurrence. In particular, neural spike trains display a wide variety of history dependent behaviors. The electrophysiological properties of neural membranes limit how fast a neuron can spike immediately after another spike. This leads to an absolute refractory period following a spike, when the neuron is physically unable to fire another spike, no matter what the stimulus, and to a relative refractory period after that, during which the neuron can fire a spike, but requires a more effective stimulus in order to do so. Additionally, many neurons exhibit bursting behavior, characterized by collections of spikes firing in rapid succession. Another example of history dependence is oscillatory spiking activity. In many systems, neurons tend to fire in specific frequency bands. For example, neurons in the CA1 region of rodent hippocampus tend to fire at particular phases of the EEG theta rhythm (Buzsaki et al., (1983)). In human patients with

Parkinson's disease, neurons exhibit a complex pattern of inhibition and excitation causing them to fire more often 10-20 msec after a previous spike (Levy et al., 2000).

In order to model these history-dependent behaviors, we could start by describing how the probability of a spike in a small time bin depends on the time since the previous spike. If a point process depends only on the previous spike time then that implies that the inter-spike intervals are independent. In this case, it might be most appropriate to model the ISI distribution rather than the spiking probability in a small time bin. However, to capture more complicated behaviors such as bursting or oscillatory spiking, it may be necessary to go back more than one spike in the past. In those cases, we will need a general framework to describe arbitrary history dependencies. We will start by looking at Poisson processes, which have no history dependence. Next we will examine renewal processes, which have the simplest form of non-trivial history dependence, that is, they depend only on the time of the most recent spike. Finally we will discuss a framework for describing general point processes that can exhibit complex patterns of history dependence, such as those observed in real physical and physiological systems.

IV. Poisson processes are point processes for which spiking probabilities do not depend on occurrence or timing of past spikes.

Perhaps the simplest class of point process model is the **Poisson process**. As we shall see, Poisson processes are characterized by the fact that the probability of firing a spike in a small time interval is independent of the firing activity at all other times. In particular, this means that Poisson processes are independent of their past spiking history. The probability models used to describe these processes make use of this independence assumption and hence the methods associated with fitting models, measuring goodness-of-fit, and making inferences about Poisson data are easy to implement and understand. For this reason, Poisson processes are often the first-order models used to characterize spiking systems. However, as we will see, most neural data is not well described as a Poisson process. However, developing the theory associated with Poisson processes will be extremely useful when generalizing to arbitrary history-dependent point processes.

All Poisson processes possess this history independence property. If we further assume that the spiking probability models do not depend on time, then we get a **simple** or **homogeneous Poisson process**. A simple Poisson process is characterized by a single rate parameter, λ. In order to develop an intuition about simple Poisson processes, we will provide three different, but equivalent definitions, in terms of different random variables.

Definition 1: A Poisson process of rate λ is a point process satisfying the following conditions:

a) For any interval, $(t, t+\Delta t)$, $\Delta N_{(t,t+\Delta t)} \sim \text{Pois}[\lambda \cdot \Delta t]$.

b) For any non-overlapping intervals, $(t, t+\Delta t)$ and $(s, s+\Delta s)$, $\Delta N_{(t,t+\Delta t)}$ and $\Delta N_{(s,s+\Delta s)}$ are independent.

The first condition states that for any time interval, the point process increment for that interval is a Poisson random variable, with parameter $\lambda \cdot \Delta t$. Remember that the

expectation of a Poisson random variable is just equal to its parameter. Here, the parameter, and hence the expected number of spikes in an interval scales linearly with the length of that interval. Notice that the distribution of each increment depends on Δt, the size of the interval, but not on t, the starting time of the interval. Since this distribution does not depend on t, we say that it is stationary. The second condition states that increments from non-overlapping intervals are independent. In other words, the distribution of the number of events in any interval does not depend on any of the spiking activity outside that interval. Therefore, an equivalent statement of this definition is that a Poisson process is a point process with stationary, independent increments.

If we discretize the observation interval into equally sized bins of size Δt, the discrete-time increments of a Poisson process are **independent, identically distributed (i.i.d.)** according to a Poisson distribution with parameter $\lambda \cdot \Delta t$. Each of the discrete time bins are disjoint, and therefore all of the increments are independent. It is this definition that gives the Poisson process its name. A stochastic process is just an indexed collection of random variables, and here we have a collection of random increments, indexed by time, each of which is Poisson.

Another important feature of Poisson processes that is evident from this definition is the fact that in any interval, the expectation of the number of spikes fired is exactly equal to the variance of the number of spikes fired. This feature will become important when we develop statistical tests to determine whether Poisson process models accurately describe the structure in observed spiking data.

There is one technical point we need to check in order to make sure that this definition is self consistent. How can we be sure that it's possible that the distribution of any sized interval is always Poisson? That is, if we have two adjacent intervals, each of which has a Poisson increment, will it necessarily be the case that the increment over both intervals, which is just the sum of those two increments, will also be Poisson? In the appendix to this chapter (Section AI), we present a simple proof that shows that this is always the case, and that the above definition is consistent.

Our second definition of a Poisson process focuses on the discrete time increments for small time steps. For most physical systems, and certainly for neural spiking data, it is possible to pick a discrete time bin size, Δt, such that at most one spike can fire in any one bin. When this is true, we say that the point process is **orderly**, since this ensures that it is always possible to tell which of two spikes occurs first and to therefore put them in chronological order.

Clearly the expected number of spikes in a bin will decrease as the length of that interval becomes small. If the length of a bin becomes sufficiently small then the probability of firing no spikes is close to one, the probability of firing a single spike is small, and the probability of firing more than one spike is negligibly small. In this case, the probability structure of the process in a sufficiently small bin becomes equivalent to just flipping a biased coin. This reasoning leads to the following definition for a Poisson process.

Definition 2: Partition the observation interval, $(0,T]$, into regular bins of size Δt, and let each ΔN_k be an independent Bernoulli$[p]$ random variable with $p = \lambda \cdot \Delta t$. Then in the

limit as $\Delta t \to 0$, the distribution of the counting process $N(t_k) = \sum_{i=1}^{k} \Delta N_i$ will approach that of the counting process for a Poisson process of rate λ.

In other words, a Poisson process is the limit of an independent Bernoulli process. This definition provides some fundamental intuition about Poisson processes and point processes in general. As the time interval we consider becomes sufficiently small so than no more than one spike can occur in any bin, we only need to characterize the probability, p, of observing a spike in that bin. The probability of not having a spike in that bin is then $1-p$, and the probability of having more than one spike in that bin is negligibly small. In this case, the Poisson distribution from Definition 1 is nearly identical to a Bernoulli distribution. For a simple, or homogeneous Poisson process, each bin has the same probability of having a spike, and that probability is entirely determined by the rate of the Poisson process and the length of each bin. This is equivalent to the stationary increment property we discussed with respect to the first definition of the Poisson process.

What is the appropriate bin size in order to ensure that at most one spike can be observed in a single bin? That depends on the process being modeled. For the spiking activity of a neuron, this assumption is biophysically plausible because neurons have an absolute refractory period (Kandel, Schwartz, and Jessel, 2000). A bin length of $\Delta t \leq 1$ millisecond is typically sufficient to ensure that at most one spike can fire in any bin.

Both of these definitions for a Poisson process relate to the distribution of the point process increments. In the appendix to this chapter (Section AII), we prove that these two definitions are equivalent. Instead of defining the Poisson process in terms of the increments process, we can equivalently define it in terms of the waiting time between occurrences by defining the **inter-spike interval (ISI) distribution**. Once we define the probability model for the increments, it implicitly defines a probability structure for any other random variables that describe the same data. Let X_i be the inter-spike interval between the $(i-1)^{\text{st}}$ and i^{th} spike times. Then the event that $X_i > t$ for some time t is equivalent to the event that $\Delta N_{(S_{i-1}, S_{i-1}+t]} = 0$. By definition 1, $\Pr(X_i > t) = \Pr(\Delta N_{(S_{i-1}, S_{i-1}+t]} = 0) = e^{-\lambda t}$. Therefore, the cumulative distribution function (CDF) for the ISIs is $F_{X_i}(t) = \Pr(X_i \leq t) = 1 - e^{-\lambda t}$, which is the CDF of an exponential random variable with mean $E[X_i] = \lambda^{-1}$. Furthermore, since the spiking probabilities of a Poisson process do not depend on the occurrences or times of past spikes, is follows that the time between two spikes does not depend on the ISI between any other spikes. This argument leads to our third equivalent definition for a Poisson process.

Def 3. A Poisson process of rate λ is a point process with i.i.d exponential ISIs with mean λ^{-1}.

The fact that they have exponential waiting times leads to another important feature of simple Poisson processes. The waiting time from any point in time until the next spike time is independent of the time since the last spike, or of any other spiking activity in the

past. This characteristic of Poisson processes is known as the **memoryless** property. Mathematically, if we let X_i be the i^{th} waiting time then the memoryless property can be written $\Pr(X_i > t) = \Pr(X_i > s+t \mid X_i > s)$. That is, if we've already waited a time s since the last spike, and haven't observed any new spikes in that time, then the probability distribution of the remaining time to wait until the next spike is the same as the original waiting time distribution right after the last spike. It is easy to see that this is the case if the X_is are i.i.d. exponential, $\Pr(X_i > t) = e^{-\lambda t}$. In this case,

$$\Pr(X_i > s+t \mid X_i > s) = \frac{\Pr(X_i > s+t, X_i > s)}{\Pr(X_i > s)} = \frac{\Pr(X_i > s+t)}{\Pr(X_i > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t}.$$

This definition reinforces the idea that the spiking distribution for Poisson processes does not depend on past spike times or occurrences. Just as the distribution of any increment does not depend on past spiking activity, so too is the distribution of the time you have to wait until the next spike. Not only is it the case that the Poisson process is memoryless, but it can be readily shown that it is the only memoryless process (Ross, 1996).

Inhomogeneous Poisson processes have time-varying firing rates.

We made two major assumptions in defining a simple Poisson process: 1) that the increments were stationary, and 2) that they were independent for non-overlapping intervals. The first step in modeling a larger class of point processes is to eliminate the stationarity assumption. In other words, we would like to construct a class of models where the probability of firing a spike in a small interval varies in time.

Perhaps the easiest place to start is with our second definition for the simple Poisson process. In that case, we assumed that for a small interval, $(t, t+\Delta t]$, the probability of observing a spike is some probability p that is directly proportional to the size of the interval, Δt, the probability of observing more than one spike in this interval is negligibly small, and the probability of observing no spikes is approximately $1-p$. This process is stationary because the probability of a single spike is a function of the size of the interval, but not of the time t.

In order to construct a binary process that is not stationary, we simply need to make the probability of firing a spike in this interval depend on t. In particular, we define a **Poisson rate function**, $\lambda(t)$, such that $\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(\Delta N_{(t,t+\Delta t]} = 1)}{\Delta t}$. This rate function defines the instantaneous probability of observing a spike at each point in time. We can now define an inhomogeneous Poisson function as the limit of a Bernoulli process, as we did in our second definition for a simple Poisson process.

Definition: Partition the observation interval, $(0, T]$, into regular bins of size Δt, and let ΔN_k be an independent Bernoulli$[p_k]$ process with $p_k = \lambda(t_k) \cdot \Delta t$. Then in the limit as

$\Delta t \to 0$, $N(t)$ will approach the counting process for an inhomogeneous Poisson process with rate function $\lambda(t)$.

Since this definition characterizes the probability of firing a spike in any small bin, it also implicitly defines the probability distribution of the number of spikes in any interval. In particular, it is not difficult to show that the number of spikes in any interval from time a to time b has a Poisson distribution with parameter $\int_b^a \lambda(t)dt$. A full proof is given in the appendix (Section AI), but the intuition behind this result is easily explained. From the above definition, we know that the probability distribution of each small increment is approximately Bernoulli, and as we discussed with simple Poisson processes, Bernoulli and Poisson distributions are nearly identical for small increments. Therefore, each small increment ΔN_k is approximately Poisson with parameter $\lambda(t_k)\Delta t$. Previously, we used the result in the appendix (Section AI) that the sum of Poisson random variables is itself Poisson whose parameter is the sum of those of the original Poisson variables. Therefore, the number of events in any large interval will have a Poisson distribution with a parameter $\sum \lambda(t_k)\Delta t$. In the limit as $\Delta t \to 0$, the sum in the parameter becomes an integral, and all of the approximations become exact. In other words, each infinitesimal increment has a Poisson distribution, which ensures that any larger increment must also be Poisson.

Since the discrete increments in the above definition of the inhomogeneous Poisson process are independent, it follows that the sums of two disjoint groups of those increments will also be independent. Therefore, we expect non-overlapping intervals to have independent Poisson increments. This leads to an alternate definition of the inhomogeneous Poisson process:

Definition: An inhomogeneous Poisson process with rate function $\lambda(t)$ is a point process satisfying the following conditions:

a) For any interval, $(t_{\text{start}}, t_{\text{end}})$, $\Delta N_{(t_{\text{start}}, t_{\text{end}})} \sim \text{Pois}\left[\int_{t_{\text{start}}}^{t_{\text{end}}} \lambda(t)dt\right]$.

b) For any non-overlapping intervals, $(t_{\text{start}}, t_{\text{end}})$ and $(s_{\text{start}}, s_{\text{end}})$, $\Delta N_{(t_{\text{start}}, t_{\text{end}})}$ and $\Delta N_{(s_{\text{start}}, s_{\text{end}})}$ are independent.

The reason this process is called an inhomogeneous Poisson process is clear. It still has Poisson increments, but each increment has its own mean, determined by the value of the rate function over the interval in question. This process no longer possesses the independent increments property, but still has independent increments. As a result, this process also still has the memoryless property, by which the probability of spiking at any instant does not depend on occurrences or timing of past spikes.

It is also possible to define an inhomogeneous Poisson process in terms of its ISI distribution. Since this distribution is nonstationary, it is useful to define it in terms of the distribution of the next spike time given the most recent spike time. Clearly, if the time of the last spike is known, then the distribution of the waiting time is equal to that of the difference between the next spike time and the previous one.

We can compute the distribution of the time to the next spike given the previous spike time by noting that the event that the time until the next spike is greater than some time s_i, i.e. $\{S_i > s_i \mid S_{i-1} = s_{i-1}\}$, is equivalent to the event that no spikes occur in the interval $(s_{i-1}, s_i]$. Therefore, $\Pr(S_i > s_i \mid S_{i-1} = s_{i-1}) = \Pr(\Delta N_{(s_{i-1},s_i]} = 0) = \exp\left\{-\int_{s_{i-1}}^{s_i} \lambda(t)dt\right\}$, and the cumulative distribution function (CDF) is $\Pr(S_i \le s_i \mid S_{i-1} = s_{i-1}) = 1 - \exp\left\{-\int_{s_{i-1}}^{s_i} \lambda(t)dt\right\}$. The probability density function of the next spike time is given by derivative of CDF, $f_{S_i}(s_i \mid S_{i-1} = s_{i-1}) = \frac{d}{dt}\left(1 - \exp\left\{-\int_{s_{i-1}}^{s_i} \lambda(t)dt\right\}\right)$, which implies,

$$\boxed{f_{S_i}(s_i \mid S_{i-1} = s_{i-1}) = \lambda(s_i)\exp\left\{-\int_{s_{i-1}}^{s_i} \lambda(t)dt\right\}}$$

As established by the above definitions, this density is independent of any spiking activity that occurs outside of the interval $(s_{i-1}, s_i]$.

V. Renewal processes are simple point processes in which the inter-spike intervals are independent.

The simple Poisson process developed above assumed that the point process increments were both stationary and independent of past spiking history. Since simple Poisson processes cannot describe spiking activity with properties that change over time, we generalized to inhomogeneous Poisson processes by eliminating the stationarity assumption, but preserving the independence assumption. Therefore, both classes of Poisson processes have no history dependence. However, as discussed in section III, most systems that produce point process data have physical mechanisms that lead to history dependent spiking, which cannot be explained with Poisson models. Therefore, it is necessary to further generalize our point process models by removing the independence assumption to capture observed history dependent structure.

The simplest type of point process with nontrivial history dependence is the renewal process. A **renewal process** is a point process for which the probability of firing a spike at any point in time can depend on the occurrence time of the last spike, but not on any spikes before then. It is therefore typical to express the probability models for renewal processes in terms of the inter-spike interval distributions. Specifically, renewal processes have independent identically distributed inter-spike intervals. We already showed in section IV that simple Poisson processes have i.i.d. exponential inter-spike intervals. Therefore, renewal processes are a generalization of the simple Poisson process allowing other ISI distributions beyond the exponential. It is important to understand that for renewal processes, each inter-spike interval is independent of every other ISI, but the probability of a spike at any point in time can still depends on past spike times.

A renewal model is specified by writing down a distribution function for the inter-spike intervals. Typically, this takes the form of a probability density function, $f_{X_i}(x_i)$, where

x_i can take values in $[0,\infty)$. In principle, we can define a renewal process using any probability distribution that takes on values between zero and infinity, however there are some classes of probability models that are more commonly used either because of their distributional properties, or because of some physical or physiological features of the underlying process.

For example, the Gamma distribution is a common basis for renewal models that is defined by two values, a mean and a shape parameter. The shape parameter gives it the flexibility to capture a number of characteristics that are often observed in point process data (Fig 3a). If this shape parameter is equal to one, then the Gamma distribution simplifies to an exponential, which as we have shown, is the ISI distribution of a simple Poisson process. Therefore, renewal models based on the Gamma distribution generalize simple Poisson processes, and can be used to address questions about whether data is actually Poisson. If the shape parameter is less than one, then the density drops off faster than an exponential. This can be useful in describing point processes that fire in rapid bursts. If the shape parameter is greater than one, then the Gamma density function takes on the value zero at $x_i = 0$, rises to a maximum value at some positive value of x_i, and then falls back to zero. This can be useful in describing **refractoriness** in point processes, that is the property that after firing a spike, the process is less likely to fire again immediately afterward. Therefore, this simple class of distributions with only two parameters is capable of capturing a number of different types of history dependent structure.

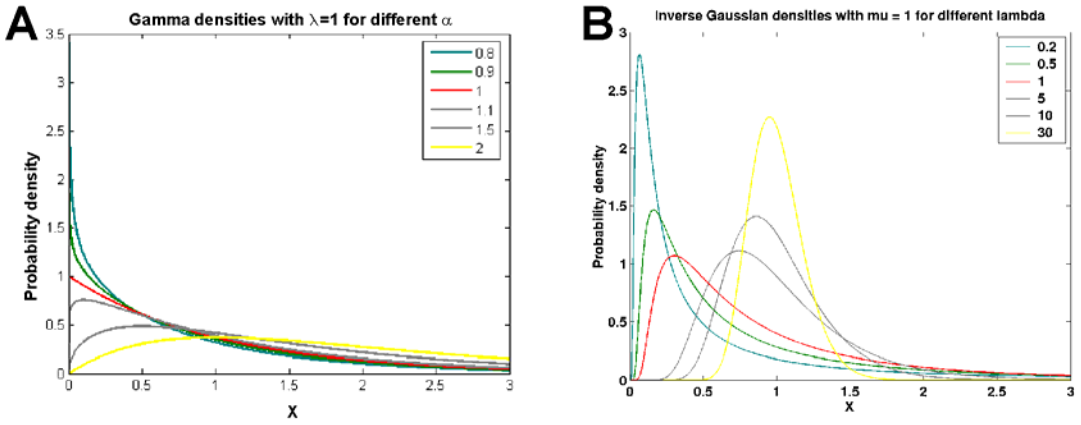

Figure 6. A) Examples of Gamma density functions with fixed mean and different shape parameters. B) Examples of inverse Gaussian density functions with fixed mean and different shape parameters.

For neural spiking data, perhaps a more appropriate distribution for renewal models is the inverse Gaussian. While the Gamma distribution is simple and flexible, it doesn't directly relate to the physiological properties of neurons. On the other hand, the inverse Gaussian renewal model is motivated by simple dynamical models of the mechanisms underlying the generation of action potentials. Specifically, if we construct a simple **integrate-and-fire** model for the neuron, wherein the membrane potential at any point in time is the integrated value of random synaptic inputs from other neurons and a spike is fired once that potential reaches a fixed threshold, then we can compute the inter-spike distribution based on the distribution of the inputs (Fig 7). If the input distribution is taken to be a simple Gaussian white noise process, then from section 2.6.5, we see that the

spiking activity turns out to be an Inverse Gaussian renewal process (Tuckwell, 1988). Like the Gamma distribution, the inverse Gaussian distribution is defined by two parameters, a mean and a shape parameter. For all parameter values, the inverse Gaussian density function is zero at $x_i = 0$, rises to a maximum value at some positive x_i, and decays back to zero (Fig 6b). The shape parameter determines how quickly the density rises and falls, and determines features of the model such as its tendency to fire in rapid bursts and its refractoriness.

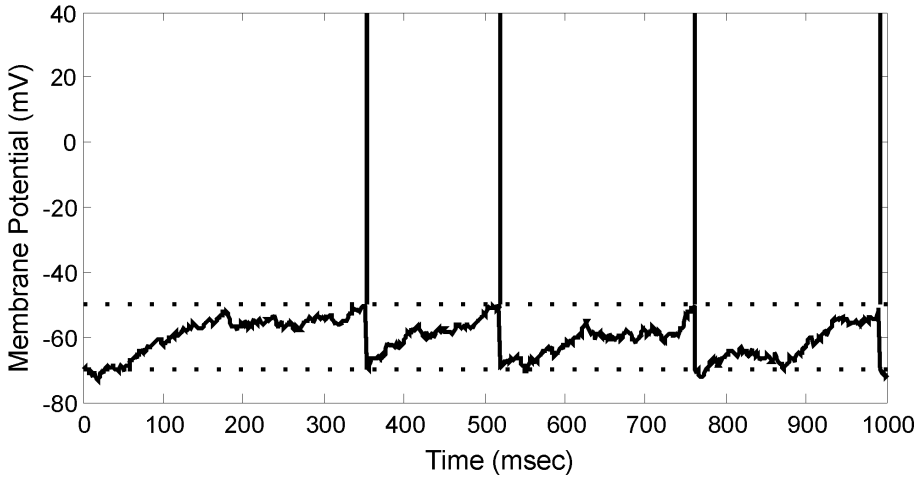

Figure 7. Example of an integrate-and-fire neuron. The membrane integrates random Gaussian inputs until it reaches a threshold potential, at which point it fires a spike and resets to the resting potential.

When discussing whether a process is stationary, it is important to specify to which collection of random variables one is referring. For simple Poisson processes, the distributions of both the point process increments and the inter-spike intervals are stationary. For renewal processes the increments are typically nonstationary and depend on the time of the last spike, but the ISIs are stationary. It is also possible to construct inhomogeneous inter-spike interval probability models, with nonstationary ISIs. For Poisson processes, we made the process inhomogeneous by replacing the constant rate parameter with a time varying rate function. For general renewal processes, we can create an inhomogeneous process by performing a time rescaling calculation (Barbieri et al., 2001). As we shall see, inhomogeneous point process models are useful for describing data from stimulus-response experiments.

VI. The conditional intensity function specifies the joint probability density of spike times for a general point process.

Many systems that produce point process data have history dependent structure that makes Poisson and renewal models inappropriate. It is useful to define a unified mathematical construct that will allow us to describe any physically relevant point process. Previously, we described the structure of an inhomogeneous Poisson process by defining a rate function that characterized the instantaneous probability of firing a spike at each instant in time. Similarly, any point process can be completely

characterized by its **conditional intensity function,** $\lambda(t \mid H_t)$ (Daley and Vere-Jones, 2003), defined as

$$\boxed{\lambda(t \mid H_t) = \lim_{\Delta t \to 0} \frac{\Pr(\Delta N_{(t,t+\Delta t]} = 1 \mid H_t)}{\Delta t},} \tag{1}$$

where $\Pr(\Delta N_{(t,t+\Delta t]} = 1 \mid H_t)$ is the instantaneous conditional probability of a spike and H_t is the history of the spiking activity up to time t. Since the probability of a spike in any interval must be non-negative, so too must be the conditional intensity function. This conditional intensity function expresses the instantaneous firing probability and implicitly defines a complete probability model for the point process. It will therefore serve as the fundamental building block for constructing the likelihoods and probability distributions needed for point process data analysis.

By construction, the conditional intensity is a history-dependent rate function because it defines a probability per unit time. We will see below that it generalizes the concept of a rate function for a Poisson process. It is important to realize that the conditional intensity can be a stochastic process itself, since it can depend on spiking history, which is stochastic. A conditional intensity functions that depends on history or on any other stochastic process is often called a **conditional intensity process**, and the resulting point process is called a **doubly stochastic point process**.

Additional important intuition behind the conditional intensity function can be gained by choosing Δt to be a small time interval and re-expressing Eq. 1 as

$$\Pr(\Delta N_{(t,t+\Delta t]} = 1 \mid H_t) \approx \lambda(t \mid H_t)\Delta t\,. \tag{2}$$

Equation 2 states that the conditional intensity function multiplied by Δt gives the probability of a spike event in a small time interval Δt. Because $\lambda(t \mid H_t)$ is defined in continuous time, we see that it defines the probability of a spike event in any small time interval.

As discussed in section IV, we typically assume that the point processes we model are orderly. That is, that for a sufficiently small interval, the probability of firing more than one spike in negligibly small compared to the probability of firing one spike. Mathematically, this is equivalent to the statement that

$$\lim_{\Delta t \to 0} \frac{\Pr(\Delta N_{(t,t+\Delta t]} > 1 \mid H_t)}{\Delta t} = 0\,. \tag{3}$$

This assumption is biophysically plausible for a point process model of a neuron because neurons have an absolute refractory period (Kandel, Jessel and Swartz, 200_). For most neurons the probability of firing more than one spike is negligibly small for $\Delta t < 1$ msec.

Together Eqs. 2 and 3 imply that given the conditional intensity function, either zero or one spike can possibly fire in a small time interval. Therefore, the probability of a spike

occurring can be analyzed as a Bernoulli or coin flipping process in which for any small interval Δt the probability of a spike is

$$\Pr(\text{spike in } [t, t+\Delta t) \mid H_t) \approx \lambda(t \mid H_t)\Delta t \tag{4}$$

and the probability of no spike is

$$\Pr(\text{no spike in } [t, t+\Delta t) \mid H_t) \approx 1 - \lambda(t \mid H_t)\Delta t\,. \tag{5}$$

This is one sense in which the conditional intensity function characterizes a spike train. In section IV, we showed that Poisson processes could be thought of as a local Bernoulli process with a spiking probability determined by the Poisson rate function. Equations 4 and 5 generalize this result and show that for orderly point process, in a small time interval, given the history up to time t, spike events are described by a local Bernoulli process. This is the reason that point processes are conceptually simple to understand. At each instant, you either observe a spike or you don't observe a spike. The probability of observing a spike can be a function of past spiking activity, as well as other stochastic or time varying signals, and is characterized by the conditional intensity.

In the appendix (Section AIII), we use this Bernoulli approximation for the increments and take the limit as $\Delta t \to 0$ to show that the probability density for the time to the next spike given the spiking history is

$$\boxed{f_{S_i}\left(s \mid H_{s_{i-1}}\right) = \lambda(s \mid H_s)\exp\left\{-\int_{s_{i-1}}^{s} \lambda(t \mid H_t)dt\right\}}, \tag{6}$$

where s_{i-1} is the observed $(i-1)^{\text{th}}$ spike time and $H_{s_{i-1}}$ is the observed spiking history up until and including time s_{i-1}. This density function also defines the ISI distribution, since the event that the next spike time occurs at a time s given the spiking history up until the last spike time, $H_{s_{i-1}}$, is equivalent to the event that the i^{th} ISI is equal to $s - s_{i-1}$.

One way to interpret Equation 6 is to look at the two terms in the product on the right hand side separately. In the derivation of equation 6 in the appendix (Section AIII), we show that the second term, $\exp\left\{-\int_{s_{i-1}}^{s} \lambda(t \mid H_t)dt\right\}$ gives the probability of firing no spikes between times s_{i-1} and s. The first term, $\lambda(s \mid H_s)$, characterizes the instantaneous distribution of firing a spike at time s. Therefore the product in Equation 6 describes the probability of not firing any spikes between s_{i-1} and s, and then firing at exactly time s.

We can take this argument one step further to show that in an observation interval $(0, T]$, the joint probability density of observing a spike train with spikes occurring at the times $s_1, \ldots, s_{N(T)}$ is

$$f_{S_1,\ldots,S_{N(T)}}\left(s_1,\ldots,s_{N(T)}\right)=\prod_{i=1}^{N(T)}\left(\lambda(s_i\mid H_{s_i})\right)\exp\left\{-\int_0^T\lambda(t\mid H_t)dt\right\}, \tag{7}$$

where $N(T)$ is the total number of spikes observed in the interval $(0,T]$, and $S_{N(T)}$ is the time of the last observed spike. The complete derivation of this result is shown in the appendix (Section AIII). As before, we can interpret Equation 7 by breaking the product on the right hand side into two terms. The $\prod_{i=1}^{N(T)}\left(\lambda(s_i\mid H_{s_i})\right)$ term characterizes the distribution of firing at exactly the observed spike times, $s_1,\ldots,s_{N(T)}$. The $\exp\left\{-\int_0^T\lambda(t\mid H_t)dt\right\}$ term gives the probability of not firing any other spikes in the observation interval $(0,T]$. Therefore, Equation 7 describes the distribution of firing only at the observed spike times, and nowhere else. This joint probability density fully characterizes a spike train, and will be the backbone for most of our statistical analyses. Thus it is clear that once we define a conditional intensity function for a point process, we can compute the probability distributions necessary for performing statistical inference.

The homogeneous and inhomogeneous Poisson processes discussed in section IV are specific cases of general point processes that are defined by specifying a particular structure on the conditional intensity function. For a homogeneous Poisson process, the probability distribution of firing is independent of time and history. In other words, a homogeneous Poisson process has a constant conditional intensity that is equal to the Poisson parameter,

$$\lambda(t\mid H_t)=\lambda_0\,. \tag{8}$$

For an inhomogeneous Poisson process, the probability of firing is given by a time-dependent rate function but is still independent of past firing history. Therefore, an inhomogeneous Poisson process has a conditional intensity that is a function of time but not of history,

$$\lambda(t\mid H_t)=\lambda(t)\,. \tag{9}$$

Thus, the conditional intensity function can be thought of as a history-dependent generalization of the Poisson rate function.

Because the spiking propensity of a neuron depends critically on history defined as the recent state of its local milieu, including its own recent spiking activity, that of the other neurons in its network and external stimuli, certainly simple Poisson models should not be the first models considered for analyzing neural spike trains. Inhomogeneous Poisson models that capture the time-dependent nature of a stimulus may be very useful for relating a neuron's spiking activity (Brown et al 1998, Brown et al. 2001, Barbieri et al. 2004; Barbieri and Brown 2006) even though it does not consider the intrinsic dynamics of the neuron and the network dynamics, if multiple neurons are recorded simultaneously.

Although the conditional intensity function is defined in continuous time, it is useful to analyze it in discrete time. In section III, we defined a discrete-time partition of an observation interval $(0,T]$ into n subintervals each of length $\Delta t = T/n$. We chose n large so that by the orderliness assumption (Eq. 3) there is at most one spike in any subinterval, and defined ΔN_k, for $k = 1,...,n$, to be the increment process, where $\Delta N_k = 1$ if there is a spike in $[k\Delta,(k+1)\Delta)$ and 0 otherwise. Similarly, we can define the conditional intensity for each interval to be its value at the beginning of the interval, $\lambda(k\Delta t \mid H_{k\Delta t})$. To simplify notation, we denote the discretized conditional intensity function as

$$\lambda_k = \lambda(k\Delta t \mid H_{k\Delta t}). \tag{10}$$

For a fine discretization, the probability of observing a spike is approximately given by, $\Pr[\Delta N_k = 1] \approx \lambda_k \Delta t$, the probability of not observing a spike is approximately $\Pr[\Delta N_k = 0] \approx 1 - \lambda_k \Delta t$, and the probability of observing more than one spike is negligibly small.

A point process model expresses the conditional intensity as a function of time, history, and other variables.

Equation 1 defines the conditional intensity function in terms of the instantaneous firing probability of the point process. However, we do not use this equation to get the conditional intensity, in general, since this probability distribution is typically unknown *a priori*. In constructing point process models, we are trying to define those exact probabilities. Therefore, rather than constructing the conditional intensity based on the firing probability, we typically write down a model for the conditional intensity, which implicitly defines the probability model for any spiking data. Thus, a conditional intensity model provides a way to express the probability model for a spiking process.

The first step in writing down a conditional intensity model for a point process is determining what factors, or **covariates**, can influence the occurrence times of that process. In section II, we showed that that spiking history often plays an important role in determining when the next spike will occur. This is one class of covariates that should be considered for most point process models. If the point process being modeled is one of a larger collection of point processes that interact with each other, it may be useful to consider the firing histories of the other point processes in the model as well. For many experiments dealing with spiking data, there are other signals, or external covariates, besides history terms that affect the point process. These external covariates are often recorded simultaneously with the point process. For example, in any stimulus-response experiment, it is expected that some function of the stimulus affects the firing probability. Once we establish which covariates can influence the spike times of the point process, we next define a model for the conditional intensity as a function of those covariates. For example, if we have a point process with a firing probability that changes as a function of time, as a function of some external covariate, $x(t)$, and is history

dependent, a conditional intensity model for that process is an expression of the form $\lambda(t \mid H_t) = g(t, x(t), H_t)$, where $g(t, x(t), H_t)$ is any nonnegative function.

For neural systems, a neuron's own spiking history, as well as the firing activity of other neurons can certainly influence the timing of future spikes. Additionally, neurons are able to represent information about biological stimuli as well as behavioral and motor outputs in the firing activity of individual neurons. The functional relation between a neuron's spiking activity and these biological and behavioral signals is often called the neuron's **receptive field**. In the examples below, we look at some simple models for specific neural receptive fields.

Example 4: Simple History Dependent Spiking Model

To illustrate the effect of spiking history on current spiking probability we define a conditional intensity model that is solely a function of recent past spiking activity,

$$\lambda_k = \exp\{\alpha_0 + \sum_{j=1}^{4} \alpha_j \Delta N_{k-j}\}. \tag{11}$$

If $\Delta t = 1$ msec, then the spiking probability, $\lambda_k \Delta t$, depends on the spike occurrences in the last 4 msec. This model has 4 covariates, ΔN_{k-1}, ΔN_{k-2}, ΔN_{k-3}, and ΔN_{k-4}, and five parameters, α_0, α_1, α_2, α_3, and α_4. If we take $\alpha_0 = \log(10)$, $\alpha_1 = -100$, $\alpha_2 = -2$, $\alpha_3 = -0.5$, and $\alpha_4 = -0.1$ we see that these values of the coefficients allow a spike train process with an absolute and relative refractory period. If at any point in time, no spikes have occurred in the past 4 msec, then the firing intensity is $\lambda_k = \exp\{\log(10)\} = 10$ spikes per second. If a spike has occurred in the past millisecond, then the firing intensity drops to around $\lambda_k = \exp\{-97.7\}$, which is negligibly small. In other words, it is virtually impossible for this process to fire a spike within one millisecond of a previous spike. If a spike occurred 2 msec previously, and no other spike is present in the 4 msec history, then the firing intensity is $\lambda_k = \exp\{\log(10) - 2\} \approx 1.35$ spikes per second. This is a substantial drop from the baseline firing rate of 10 spikes per second, but not negligibly small as it was immediately after a spike. As a spike recedes into the past, its inhibitory effect on the current spiking activity diminishes. We say that this neuron has an absolute refractory period of 1 msec, when it cannot fire, and a relative refractory period of 4 msec, when the probability of firing is decreased. Under this model, if we had one spike 2 msec in the past and another spike 4 msec in the past, then the inhibitory effects of each past spike combine and $\lambda_k = \exp\{\log(10) - 2 - .1\} \approx 1.22$ spikes per second, less than the intensity caused by either past spike individually. This simple example shows that the precise timing of the previous spiking activity can alter current spiking propensity and that this can be modeled with a conditional intensity function.

Example 1. (cont'd) Conditional Intensity Model for a Retinal Neuron Under Constant Light and Environmental Conditions

Figure 2, shows the spontaneous spiking activity of a retinal neuron grown in culture and maintained under constant light and environmental conditions. This preparation ensures

that there is no explicit stimulus present, and therefore the only covariate upon which to model the activity of this neuron is its own firing history.

We can construct a history-dependent conditional intensity model similar to Eq. 11, but with more parameters stretching back further into the neuron's past history. Examining the ISI histogram in Fig. 2b, we see interesting structure going back as far as 120 msec. Being conservative, we therefore construct a conditional intensity of the form

$$\lambda_k = \exp\{\alpha_0 + \sum_{j=1}^{120} \alpha_j \Delta N_{k-j}\}. \quad (12)$$

This model has 120 history related covariates, each indicating whether or not a spike was fired in a 1 ms interval at a different time lag, and 121 parameters, giving the background intensity in the absence of past spiking and the modulation due to each of these covariates, respectively.

If we plug in the model from Eq 12 and the data shown in Figure 2c into joint spiking density in Eq. 7, we get an expression for the likelihood of the observed spiking as a function of the parameters. We will describe later how to use this likelihood to construct estimators and confidence bounds for the parameters that generated the data. In this case, the estimate for $\hat{\alpha}_0 = 3.8$, so that if there is no spikes is the past 120 msec, the conditional intensity is $\lambda_k = \exp(3.8) = 45$ spikes per second. Figure 8 shows estimates and confidence intervals for $\exp\{\alpha_i\}$ for each of the remaining model parameters. The parameters related to 0-2 msec after a spike are large negative numbers, so that $\exp\{\alpha_i\}$ is close to zero, leading to a refractory period when the neuron is much less likely to fire immediately after another spike. However, the parameters related to 4-13 ms after a spike are significantly positive leading to an increase in the firing probability. For example, if the only spike in the 120 msec history occurred 6 msec in the past, then the background conditional intensity of 45 spikes per seconds is multiplied by a factor of about 3.1, leading to a conditional intensity of 140 spikes per second. This phenomenon accounts for the rapid bursts of spikes observed in the data. There is perhaps another region of increased firing probability around 100 ms after a spike.

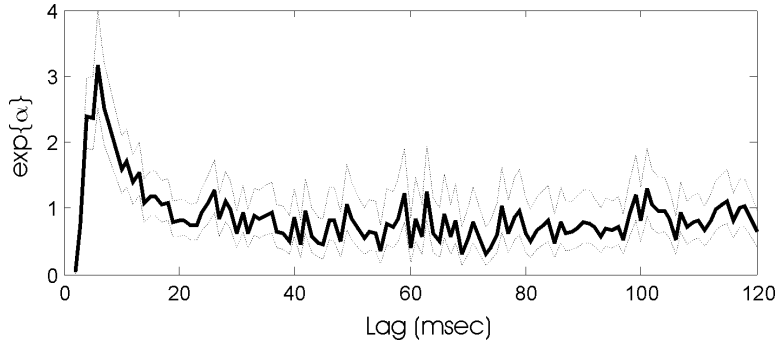

Figure 8. Parameter estimates and confidence intervals for history dependent retinal conditional intensity model.

Many of the remaining parameters are close to zero, and hence $\exp\{\alpha_i\}$ is close to one, indicating that the corresponding history term has no effect on the current spiking probability. Later, we will see that since many of these parameters do not lead to significant modulation of the firing activity, it is possible to construct a simpler model with considerably fewer parameters that describes the spiking activity as well or better than this one.

Example 2. (cont'd) Conditional Intensity Model for an M1 Neuron

As discussed earlier, primate motor cortical neurons have velocity modulated cosine tuned receptive fields. Figure 9A shows an example of the spiking activity of a neuron in primate motor cortex as a function of hand speed and direction using an occupancy normalized histogram. The neuron fires most intensely when the hand moves in a direction about 1.6 radians from east and increases as a function of hand speed.

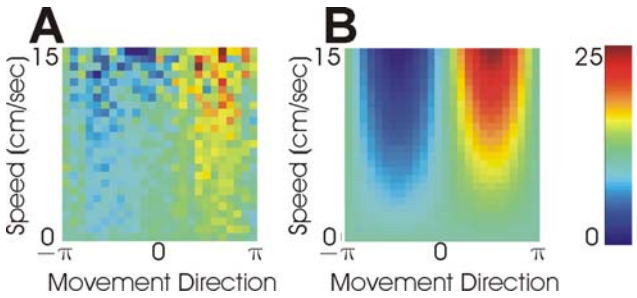

Figure 9. Example of motor cortical neural activity during a two-dimensional reaching task as a function of hand direction and speed. A) Empirical visualization of motor cortical spiking activity via occupancy normalized histogram. B) Poisson model for this spiking activity, with parameters fit by maximum likelihood.

Specifying a point process model for this neuron comes down to writing an expression for its conditional intensity in terms of the direction and speed of the hand movement. One possible model form is

$$\lambda(t) = \exp\left\{\alpha + \beta\left|v(t+150\text{ms})\right|\cos\left(\phi(t+150\text{ms}) - \phi_{\text{p}}\right)\right\} \tag{13}$$

where the model covariates are $v(t)$ and $\phi(t)$, the speed and direction of the intended hand movement, and the 150 ms lag relates the current spiking activity to movements that will occur 150 ms later. The parameters of this model are (α, β, ϕ_p), where $\exp\alpha$ is the baseline firing rate, β is the depth of modulation, and ϕ_p is the preferred direction of the neuron.

If we observe a spike train from this neuron and the corresponding subsequent hand movement, we can plug the data and the model of Equation 13 into the joint probability in Equation 7 to get an expression for the likelihood of the observed spiking as a function of the parameters. We can then find the parameters that maximize this likelihood function. Figure 9B illustrates the maximum likelihood model fit for this neuron. The preferred direction and degree of modulation as a function of speed are in good agreement with those observed in the occupancy normalized histogram.

The model of Eq 13 is an inhomogeneous Poisson model, since the conditional intensity does not depend on the past spiking history. This is unrealistic, since these neurons have been shown to have history dependent structure such as refractoriness and bursting behavior. One step toward making this model more realistic is to add a separate history component along the lines of the previous examples, to obtain a model of the form,

$$\lambda(t) = \exp\left\{\alpha + \beta\left|v(t+150\text{ms})\right|\cos\left(\phi(t+150\text{ms}) - \phi_{\text{p}}\right) + \sum_{j=1}^{120}\alpha_j \Delta N_{k-j}\right\} \tag{14}$$

where n determines how far back in the past the spiking history can affect the current spiking activity.

Example 3. (Cont'd) Conditional Intensity Model for a Hippocampal Place Cell

Hippocampal place cells have firing patterns that relate to an animal's location within an environment. Therefore a place field model should describe the conditional intensity as a function of the animal's location at each point in time. Figure 10A again shows the spiking activity of a place cell that fires maximally at a point southwest of the center of a circular environment.

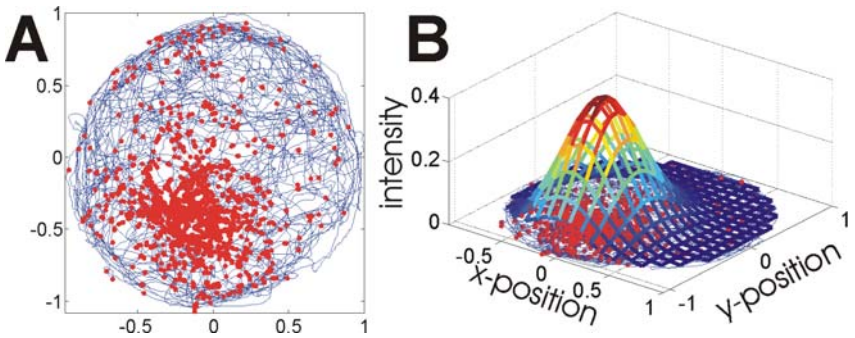

Figure 10. Spiking activity of a rat Hippocampal place cell during a free-foraging task in a circular environment. A) Visualization of animal's path (blue) and locations of spikes (red). B) Gaussian place field model for this neuron with parameters fit by maximum likelihood.

Place fields of this type have been successfully modeled with conditional intensity models that have a Gaussian shape with respect to position. For example, we can construct a conditional intensity model of the form,

$$\lambda(t) = \exp\left\{\alpha - \frac{1}{2}\begin{pmatrix} x(t)-\mu_x & y(t)-\mu_y \end{pmatrix}\begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}^{-1}\begin{pmatrix} x(t)-\mu_x \\ y(t)-\mu_y \end{pmatrix}\right\}. \tag{15}$$

The covariates for this model are $x(t)$ and $y(t)$, the animal's x and y-position. The model parameters are $(\alpha, \mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy})$, where (μ_x, μ_y) is the center of the place field, $\exp\alpha$ is the maximum firing intensity at that point, and σ_x^2, σ_y^2, and σ_{xy} express

how the intensity drops off away from the center. It is important to note that it is the shape of the place field that is Gaussian, not the distribution of the spiking activity, which is a point process.

If we observe the animal's location and the spiking activity of a Hippocampal place field, we can plug the conditional intensity model in Equation 15 and the observed spikes into the joint probability in Equation 7, to get the data likelihood as a function of the model parameters. We can then find the parameters that maximize this likelihood function. The maximum likelihood fit for the place field shown in Figure 10 is illustrated in panel B. As in the previous example, this is an inhomogeneous Poisson model that captures the spatially specific structure in the firing activity but not the history dependent structure. Later, we will discuss methods to measure goodness-of-fit between a conditional intensity model and spiking data, and construct improved models.

Defining the Conditional Intensity Function Defines the Inter-spike Interval Distribution and Vice-Versa

Equation 6 shows that given the conditional intensity function for a point process, it is simple to compute the probability density of the next spike time, given the history up until the previous spike time, which is equivalent to the ISI density function. Here we show that given the ISI density it is also straightforward to compute the conditional intensity function. Therefore, there is a one-to-one relationship between the conditional intensity function and the ISI distribution, and it is always possible to obtain the one from the other.

The key insight to establishing this relationship comes from the fact that the probability of firing no spikes in an interval can be easily computed from either of the conditional intensity or ISI density functions. In the derivation of Equation 6 in the appendix (Section AIII), we showed that

$$\Pr\left[\text{no spikes in } (s_{i-1}, s]\right] = \exp\left\{-\int_{s_{i-1}}^{s} \lambda(t \mid H_t)dt\right\}. \tag{16}$$

This is in fact the second term that appears in the product on the right hand side of Equation 6. We could instead compute this probability from the ISI density by noting that the cumulative distribution function for the ISI, which defines the probability of the ISI being less than some value s, or equivalently the probability of firing at least one spike between the last spike time, s_{i-1}, and time s, is just the integral of the density function, $F_{S_i}(s \mid H_{s_{i-1}}) = \int_{s_{i-1}}^{s} f_{S_i}(t \mid H_{s_{i-1}})dt$. Therefore, the probability of firing no spikes in this interval is

$$\Pr\left[\text{no spikes in } (s_{i-1}, s]\right] = 1 - F_{S_i}(s \mid H_{s_{i-1}}) = 1 - \int_{s_{i-1}}^{s} f_{S_i}(t \mid H_{s_{i-1}})dt. \tag{17}$$

Combining Equations (16) and (17), we see that $\exp\left\{-\int_{s_{i-1}}^{s} \lambda(t \mid H_t)dt\right\} = 1 - \int_{s_{i-1}}^{s} f_{S_i}(t \mid H_{s_{i-1}})dt$. We can now replace the second term on

the right hand side of Equation 6 to get $f_{S_i}\left(s \mid H_{s_{i-1}}\right) = \lambda(s \mid H_s)\left(1 - \int_{s_{i-1}}^{s} f_{S_i}(t \mid H_{s_{i-1}})dt\right)$. Solving for $\lambda(s \mid H_s)$ we obtain an expression for the conditional intensity function as a function of the ISI density,

$$\lambda(s \mid H_s) = \frac{f_{S_i}(s \mid H_{s_{i-1}})}{1 - \int_{s_{i-1}}^{s} f_{S_i}(t \mid H_{s_{i-1}})dt}. \tag{18}$$

Therefore, we can define a point process model by specifying a functional relationship between the covariates that influence spiking activity and either the conditional intensity, leading to a model of the form $\lambda(t \mid H_t) = g(t, x(t), H_t)$, or the ISI density, leading to a model of the form, $f_{S_i}(s \mid H_{s_{i-1}}) = h(t, x(t), H_{s_{i-1}})$. Here, $g(t, x(t), H_t)$ can be any nonnegative function, but $h(t, x(t), H_{s_{i-1}})$ must be a density function and therefore must both be nonnegative and its integral over all future time must be equal to one. Once one of these model forms is specified, any probability distributions associated with the point process can be computed. In practice, although it is possible to define any point process equivalently in terms of either a conditional intensity model or an ISI model, conditional intensity models are more commonly used when describing history dependence that goes beyond a simple renewal process.

We can gain some intuition into Equation 18 from the related field of survival analysis. In survival analysis, the objective is to estimate the rate of deaths or failures at a time s, given that a patient survived or that a machine component worked up to that time (Kalbfleisch and Prentice, 1981). This conditional failure rate is called the **hazard function**. The numerator on the right hand side of equation 18 is the probability density of an ISI of length $s - s_{i-1}$, that is the probability of no spike from time s_{i-1} up to time s, and then a spike at exactly time s. As shown by Equation 17, the denominator on the right hand of Equation 18 is simply the probability of no spike from time s_{i-1} up to time s. Therefore, we can rewrite Equation 18 as

$$\lambda(s \mid H_s) = \frac{f(\text{no spike in } (s_i, s], \text{ spike at } s)}{\Pr(\text{no spike in } (s_i, s])}, \tag{19}$$

which means that $\lambda(s \mid H_s) = f(\text{spike at } s \mid \text{no spike in } (s_i, s])$, by the definition of a conditional probability. Therefore, we see that the conditional intensity function is equivalent to the hazard function.

Example 1. (cont'd) ISI Probability Density for a Retinal Neuron

Equation 12 gives a conditional intensity model for the spontaneous spiking activity of a retinal neuron held under constant light and environmental conditions. At each spike time, we can plug this model into Eq. 18 to compute the ISI probability density for the time to the next spike. This will differ at each spike time, due to differences in past history. However, we can examine the effect of the most recent spike on its own by

taking the ISI density at each spike time, and average them together to eliminate the effect of all spikes except the most recent one. Figure 11 shows the result of this analysis, compared to the observed ISI histogram and to some simple renewal models.

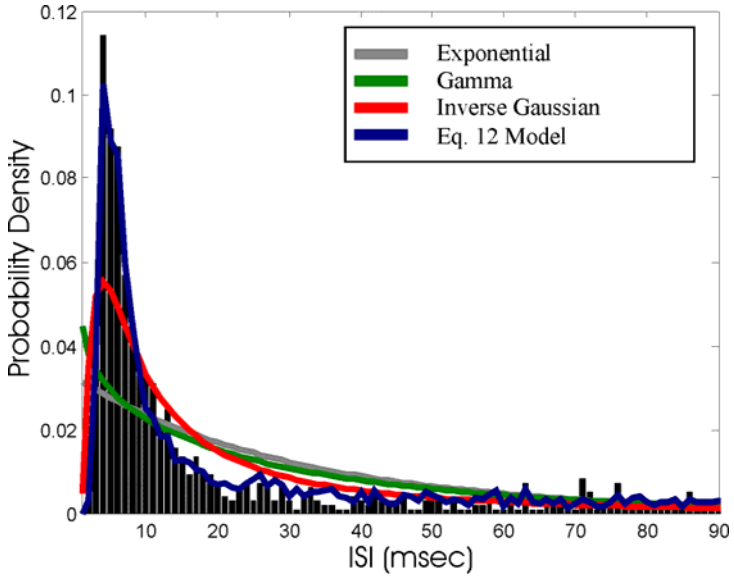

Figure 11. ISI histogram and model probability densities for exponential, gamma, and inverse Gaussian renewal models compared to conditional intensity model of Eq. 12.

The exponential and gamma renewal models both overestimates the number of very short ISIs (0-4 msec), and all three renewal models underestimate the number of ISIs between 5-10 msec and overestimate the number of ISIs between 10-60 msec. In contrast, the conditional intensity model of Eq. 12 accurately predicts the number of ISIs in all of these periods.

VII. A general temporal point process may be transformed to a constant-intensity Poisson process by rescaling time.

We can use the joint probability density function in Equation 7 to fit models by constructing data likelihoods and estimating parameter values and confidence intervals. However, once we finish the model-fitting component of our data analysis, we still need to develop methods to use those fit models to make inferences about the data and perform statistical goodness-of-fit analyses. In order to accomplish this, we need to construct statistics, which are functions of our point process data, whose distributions we can compute.

Typically, statistics are built from data samples that are independent and identically distributed. In that case, it is often possible to compute or approximate the distributions of these statistics, which because of the Central Limit Theorem tend to have Gaussian distributions. However, general point processes have complicated, nonstationary, history dependent probability models. Common statistics constructed directly from point process data often have distributions that are highly non-Gaussian, and cannot be computed analytically. In order to construct useful point process statistics, we need a

procedure to transform our point process data such that they become samples from i.i.d. random variables. One approach to this is given by the time rescaling theorem.

Time Rescaling Theorem. Given a point process with conditional intensity function $\lambda(t \mid H_t)$ and with occurrence times $0 < s_1 < s_2, ..., < s_{N(T)} \leq T$, define

$$z_1 = \int_0^{s_1} \lambda(t \mid H_t)dt \text{, and}$$

$$z_j = \int_{s_{j-1}}^{s_j} \lambda(t \mid H_t)dt \text{ , for } j = 2, ..., N(T). \tag{20}$$

Then these z_j are independent, exponential random variables with rate parameter 1.

This result comes about in a surprisingly simple manner, using the standard change-of-variables formula from probability theory. It can be seen intuitively for a single z_j by combining equations 16, 17, and 20 to show that $F_{S_i}(t \mid H_{s_{i-1}}) = 1 - \exp\left\{-\int_{s_{i-1}}^{s} \lambda(t \mid H_t)dt\right\} = 1 - \exp\{-z\}$, which is the CDF of an exponential random variable with rate parameter equal to 1. The complete proof is provided in the appendix (Section AIV).

This result is called the time rescaling theorem because we can think of the transformation as stretching and shrinking the time axis based on the value of the conditional intensity function. If $\lambda(t \mid H_t)$ is constant and equal to one everywhere, then this is a simple Poisson process with independent, exponential ISIs, and time does not need to be rescaled. Anytime when $\lambda(t \mid H_t)$ is less than one, the z_j's accumulate slowly and represent a shrinking of time, so that distant spike times are brought closer together. Likewise, anytime when $\lambda(t \mid H_t)$ is greater than one, the z_j's accumulate more rapidly and represent a stretching of time, so that nearby spikes are drawn further apart.

An illustration of the time rescaling theorem is shown in Figure 12. The upper-panel shows a Poisson rate function $\lambda(t)$ in time and a set of spike times. The process spikes when $\lambda(t)$ is high and doesn't when it is close to zero. The lower panel shows the integrated rate $\int_0^t \lambda(u)du$. The original ISIs are nonstationary and very irregular due to the time varying rate function. w_1 and w_3 are long because they contain significant portions where the rate function is close to zero, while w_2 is short because the intensity is high in this interval. If we reflect the original spike times through this function, then the rescaled spike times are more regular, and only show the degree of variability expected from i.i.d. exponential samples.

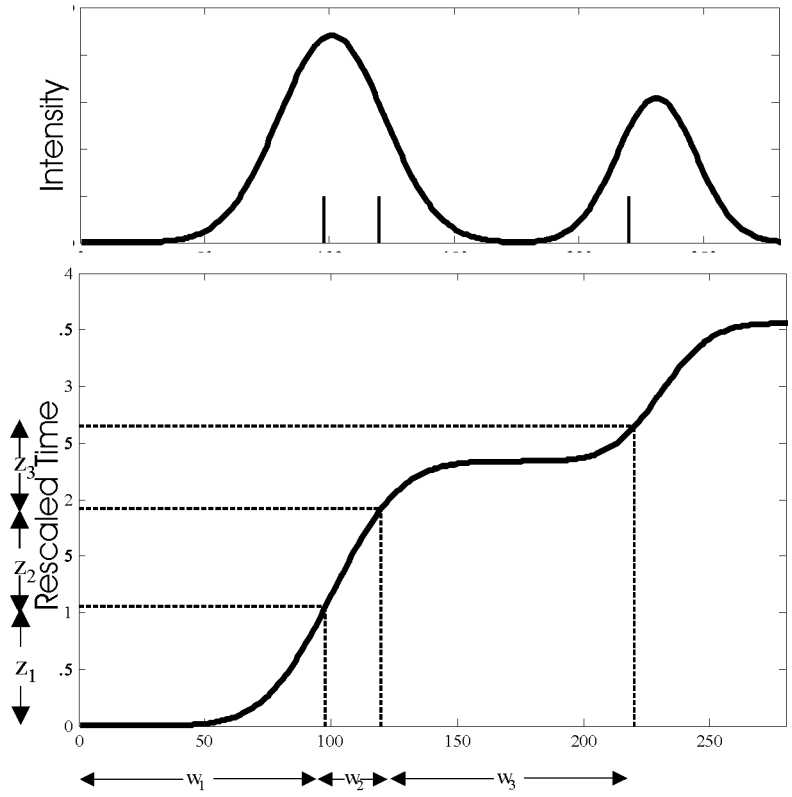

Figure 12. Example of time-rescaling for an inhomogeneous Poisson process.

In terms of statistical analysis, this result means that we do not need to compute the distributions associated with functions of dependent, nonstationary data. Instead, we can use a fit conditional intensity model to transform the data into i.i.d. samples, and then work with statistics of this transformed data, which have relatively simple distributions. For example, a statistic based on the mean of the observed ISIs will generally not have an analytically tractable distribution, but a statistic based on the mean of the rescaled ISIs will be equivalent to the mean of exponentials, which will tend to a Gaussian distribution by the Central Limit Theorem. Furthermore, we can transform the rescaled ISIs to any other i.i.d. random variables using an additional change-of-variables. In sections VIII and IX, we will use the time-rescaling theorem to visualize goodness-of-fit between a model and spiking data and to simulate spiking activity. In later chapters, we will discuss in detail how time rescaling is used to construct quantitative measures of goodness-of-fit.

As a technical point, it is important to note that we have not eliminated history-dependence by transforming the process. The spiking history is present in the realization of the conditional intensity process that is used to perform the time-rescaling calculation. Intuitively, we have sequestered the time-varying and history-dependent structure of the point process into the conditional intensity and the rescaling procedure. If the conditional intensity model accurately captures the structure of the history dependence and nonstationarity, then the rescaled z_j's will be independent and follow a simple exponential distribution.

Example 1 (cont'd). Transformed ISI's of a Retinal Neuron

We can rescale the ISIs of the spontaneous firing activity of the retinal neuron under constant light and environmental conditions using the conditional intensity model of Eq. 12. Figure 13A shows a histogram of the original ISIs for this data. The smallest bin (0-2 ms) is empty due to the refractory period of the neuron. We can also observe two distinct peaks at around 10 and 100 msec respectively. It is clear that this pattern of ISIs is not described well by an exponential distribution, and therefore the original process cannot be accurately modeled as a simple Poisson process. However the histogram in Figure 13B, which shows the result of transforming the observed ISIs according to the conditional intensity model of Eq. 12 with the model parameters shown in figure 8, is in close agreement with an exponential probability density function with mean 1. This suggests that the transformed ISIs under the model conditional intensity are close to what we would expect for the transformed ISIs under the conditional intensity that generated the data.

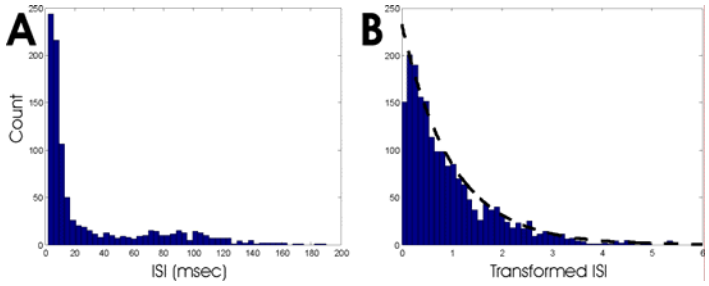

Figure 13. Histograms of A) original and B) transformed ISIs for retinal data. Dashed line in panel B is the theoretical exponential(1) probability density function.

Example 3 (cont'd). Transformed ISI's for a Hippocampal Neuron

Figure 14A shows a histogram of the actual ISIs from a hippocampal place cell of a rat during a free foraging task on a circular track. Since this spiking activity is associated with the animal's position at each point in time, we would not expect these ISIs to resemble samples from a stationary exponential distribution. Figure 14B shows the transformed ISIs using the model of Eq. 15 and the parameters that give rise to the place field model illustrated in Fig. 10B.

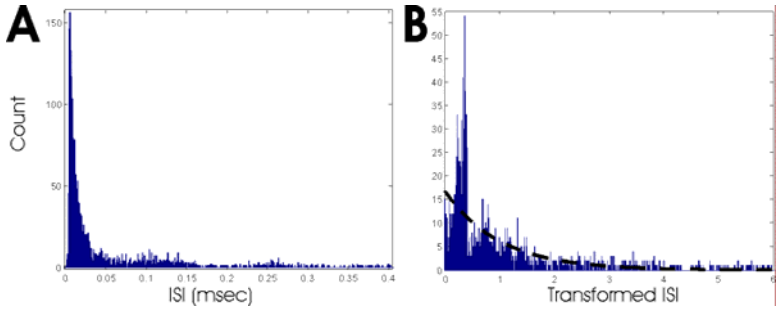

Figure 14. Histograms of A) original and B) transformed ISIs for Hippocampal place cell data. Dashed line in panel B is the theoretical exponential(1) probability density function.

While this transformation puts the bulk of the ISIs in the region that we would expect for samples from an exponential with parameter 1, we can see significant deviations from this expected distribution. When the ISIs are rescaled according to this model, we obtain fewer very short ISIs (0-0.2) than expected and many more ISIs between 0.25-0.5. This suggests that the model in Eq. 15 does not completely capture the structure in the observed spiking activity. One reason we may be failing to accurately describe the smaller ISIs is that this is an inhomogeneous Poisson model that ignores the effect of the firing history on the imminent spiking activity.

VIII. The time-rescaling method leads to Q-Q and KS plots for point processes.

One of the most important components of any statistical modeling analysis is to verify that the model accurately describes the structure observed in the data. In subsequent chapters, we will examine multiple goodness-of-fit measures between neural spiking data and conditional intensity models. Here, we discuss a pair of visualization procedures based on the time rescaling theorem that allow us to quickly envision how well a conditional intensity model describes the structure in an observed spike train and which, if any, inter-spike intervals are not well described by the model.

As discussed in section VII, for any conditional intensity model, $\lambda(t \mid H_t)$ and any set of spiking observations, $0 < s_1 < s_2, ..., < s_{N(T)} \leq T$, it is possible to construct a set of rescaled ISIs, z_j, for $j = 1, ..., N(T)$, such that if $\lambda(t \mid H_t)$ correctly describes the conditional intensity of the process generating the spike times then the z_j's will be samples from an i.i.d. exponential distribution with rate parameter 1.

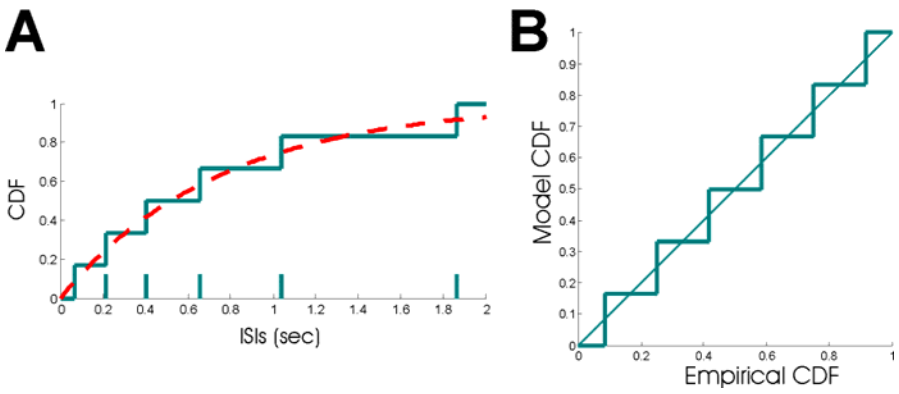

Figure 15. A) Empirical (blue) and model (red) CDFs for sample rescaled point process data. B) KS plot of rescaled ISIs closely follows 45 degree line.

We can construct an empirical cumulative distribution function, $F_{Z_j}(z)$, for these rescaled ISIs by computing for each z, the fraction of z_j's that are smaller than z. Figure 15A illustrates an example of an empirical distribution function for rescaled spiking data. The rescaled ISI data are shown as blue tics on the x-axis. The blue line is the empirical CDF for this data, which is plotted alongside the theoretical or model CDF for an exponential with rate parameter 1. We see that the empirical CDF closely follows the model CDF, indicating that the rescaled data is well described by the

exponential, and hence the conditional intensity model is correctly capturing the statistical structure of the observed point process.

A **Kolmogorov-Smirnov (KS) plot** is a plot of the empirical CDF against the model CDF, as shown in Figure 15B. If the conditional intensity model accurately describes the observed spiking data, then the empirical and model CDFs should roughly coincide and the KS plot should follow a 45-degree line. If the conditional intensity model fails to account for some aspect of the spiking behavior, then that lack of fit will be reflected in the KS plot as a significant deviation from the 45-degree line. In later chapters, we will quantify this procedure and develop a KS test for goodness-of-fit. In this example, the KS plot closely follows the 45-degree line, indicating overall agreement between the model and data. The large step-like shape of the KS plot comes from the fact that we have used very few data points. As the number of data points increases, these step shapes will become smaller.

By inverting the empirical and model cumulative distribution functions, we obtain empirical and model quantiles, respectively. For any probability level p, a quantile gives the data value such that the probability of observing a data point smaller than that value is p. Visually, the quantiles are obtained by flipping the CDFs across the 45-degree line. Figure 16A shows the empirical and model quantiles for this example.

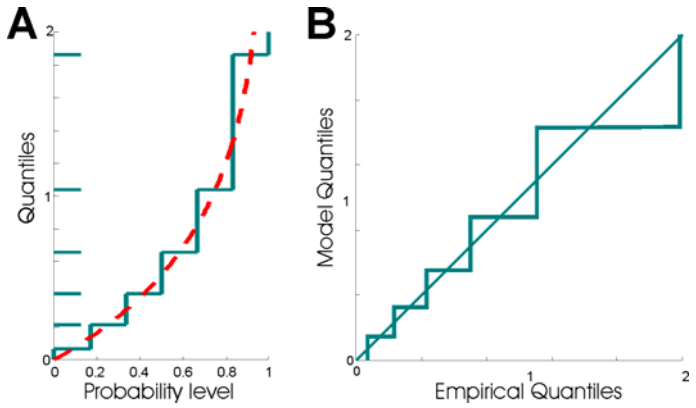

Figure 16. A) Empirical (blue) and model (red) quantiles for sample rescaled point process data. B) Q-Q plot of rescaled ISIs indicates that all quantiles are well fit.

A **quantile-quantile (Q-Q) plot** is a plot of the empirical quantiles against the model quantiles. Like the KS plot, a Q-Q plot should follow a 45-degree line if the conditional intensity function accurately describes the spiking observations. Wherever the Q-Q plot deviates from the 45-degree line, the rescaled ISIs at those values are not well fit by the conditional intensity model. As we shall see in more detail later, KS plots are most useful in determining whether or not a model accurately describes the structure in the data, while Q-Q plots are useful in determining which rescaled ISIs are inappropriately modeled.

Example 1 (cont'd). KS and Q-Q Plots of a Retinal Neuron

Having transformed the ISIs for the spontaneous activity in the retinal neuron according to the model in Eq. 12 and the parameters in Fig. 8, we con construct KS and Q-Q plots

from the empirical distributions of the rescaled intervals. Figure 17A shows the resulting KS plot, along with 95% confidence bounds, and Figure 17B shows the Q-Q plot, compared to a 45 degree line. The KS plot is close to the 45 degree line and is always inside the confidence bounds (we will discuss the construction of these bounds later). The Q-Q plot shows close agreement with the 45 degree line for small values of the rescaled intervals, and only deviates toward the tail of the distribution. This suggests that the model in Eq. 12 and the estimated parameters are able to accurately describe the observed spiking activity.

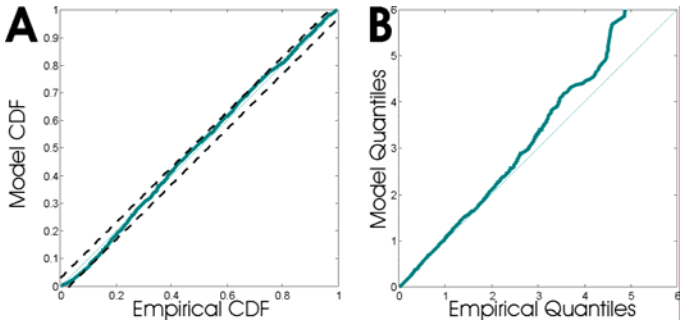

Figure 17. A) KS plot and B) Q-Q plot for distribution of rescaled intervals shown in Fig 13B.

Example 3 (cont'd). KS and Q-Q Plots of a Hippocampal Neuron

We can similarly construct KS and Q-Q plots for the transformed place cell data shown in Figure 14. In this case, the KS plot, shown in Fig. 18A deviates significantly from the 45 degree line at multiple locations. By examining the Q-Q plot in Fig. 18B, we see that the model results in too few small rescaled ISIs, too many mid-range ISIs and too few large ISIs. Again, this suggests that this inhomogeneous Poisson model for the spiking activity is unable to completely describe the structure in the data. It is likely that a similar model that incorporated spike history would provide a much closer fit to the data.

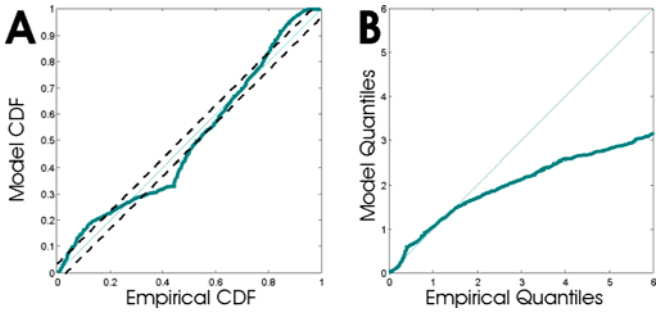

Figure 18. A) KS plot and B) Q-Q plot for distribution of rescaled intervals shown in Fig 14B.

IX. Point process theory leads to efficient methods for simulating spike data.

The previous sections have developed useful tools for modeling and analyzing observed spike data. It is also important to be able to generate new spike trains from point process models. Simulations can help us make inferences about the underlying system, can be used to compute confidence bounds about estimated quantities, and are useful in determining emergent phenomena from ensembles of interacting point processes.

Simple Poisson processes and other renewal processes are fairly simple to simulate. Since the distributions of the ISIs are independent and identically distributed, we just need to generate a sequence of independent random samples for the ISIs according to the renewal distribution. The n^{th} spike time will then be given by the sum of the first n ISIs. Most statistical software packages have built in functions to generate samples from standard distributions. In particular, to generate a homogeneous Poisson process with rate λ, we can sample the ISIs, x_i, from an exponential distribution with rate parameter λ, and take the n^{th} spike time to be $s_n = \sum_{i=1}^{n} x_i$.

However, it is not as simple a matter to simulate a general point process, since the probability of spiking changes at each instant and can depend on past history. It is usually quite difficult to simulate nonstationary mutually dependent processes directly. One simple approach, based on the Bernoulli approximation we discussed in section VI, involves partitioning the simulation interval into small bins of size Δt, and in each interval generating a spike with probability $\lambda(t \mid H_t)\Delta t$, where H_t is the history of previously generated spikes. This works well for small simulation intervals. However, as the simulation interval becomes large and as Δt becomes small, the number of Bernoulli samples that needs to be generated becomes very large, and most of those samples will be zero, since $\lambda(t \mid H_t)\Delta t$ is small. This is a very inefficient method for simulating long spike trains. Instead, we can develop alternate approaches that require a relatively small number of i.i.d. samples, but manipulate those samples so that their final distribution matches those of a general point process.

A Point process can be simulated by thinning

The thinning algorithm was originally developed by Lewis and Shedler for simulating an inhomogeneous Poisson. Ogata later developed an extension of this algorithm to the general point process using the conditional intensity function. In order to apply the algorithm, the conditional intensity function, $\lambda(t \mid H_t)$, must be bound by some constant, $\lambda_{\max}$. The algorithm follows a two-stage process. In the first stage, a set of candidate spikes is generated as a simple Poisson process with a rate $\lambda_{\max}$. This set of candidate spikes occurs more frequently than the point process we want to simulate, since $\lambda_{\max} \geq \lambda(t \mid H_t)$ at all times. The next stage involves thinning out these candidate spikes by stochastically throwing out some spikes and accepting the rest. If we accept each spike with probability $p = \lambda(s_i \mid H_{s_i}) \big/ \lambda_{\max}$, then the probability of generating and accepting a spike in a small interval $(t, t+\Delta t]$ is equal to the probability of accepting a spike given that one was generated multiplied by the probability of generating a spike in the first place, which is $\Pr(\text{generate and accept}) \approx \frac{\lambda(t \mid H_t)}{\lambda_{\max}}\left(\lambda_{\max}\Delta t\right) = \lambda(t \mid H_t)\Delta t$.

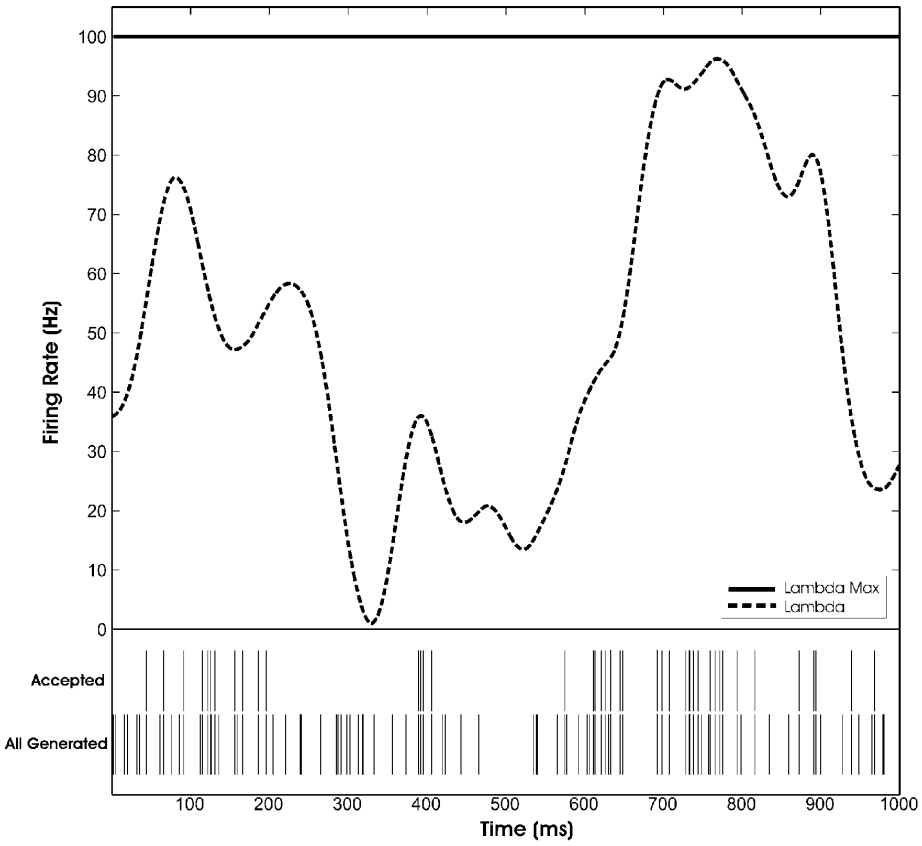

Figure 15. Simulation of an inhomogeneous Poisson process using thinning.

Figure 15 shows a simulation of an inhomogeneous Poisson process generated by thinning. The Poisson rate function, $\lambda(t)$ (dashed line), is bounded by $\lambda_{\max}$ (solid line). The bottom spike train is a simple Poisson process with rate $\lambda_{\max}$. The upper spike train has been thinned so that it represents a sample from an inhomogeneous Poisson with the shown rate function. When $\lambda(t)$ is close to $\lambda_{\max}$, many of the spikes are accepted. When $\lambda(t)$ is close to zero, most of the spikes are rejected.

In practice, thinning is typically only used when simulating inhomogenous Poisson processes with bounded rate functions. A step-by-step description of the thinning algorithm is given below.

Thinning Algorithm.

1. Initialize $i=1$ and $j=1$.
2. Sample x_i from an exponential random variable with mean $\lambda_{\max}$.
3. Set $u_i = \sum_{j=1}^{i} x_j$. Set $p = \lambda(t \mid H_t)/\lambda_{\max}$.
4. Sample b_i from a Bernoulli$[p]$ distribution.
5. If $b_i = 1$, set $s_j = u_i$ and set $j = j+1$
5. Set $i = i+1$.
6. Go to step 2.

A point process can be simulated by time-rescaling

Another approach to simulating general point processes is based on the time-rescaling theorem. Previously, we used the time rescaling theorem to transform ISIs from general, history-dependent point processes to simple Poisson processes. The theorem can also be used to convert simple Poisson process sample, which are easy to simulate, into general point processes.

Looking back at Figure 12, we can see that if we simulate a set of z_i's from an exponential distribution with rate parameter 1, as illustrated on the y-axis of the bottom panel, then we can find the times s_i that solve $z_i = \int_{s_{i-1}}^{s_i} \lambda(t \mid H_t)dt$ to obtain a spike train with the desired conditional intensity function. A step-by-step algorithm for simulating general point processes by time rescaling is given below.

Time-Rescaling Algorithm. Given an interval $(0,T]$ the simulation algorithm proceeds as follows:

1. Initialize $s_0 = 0$ and $i=1$.
2. Sample z_i from an exponential random variable with mean 1.
3. Find s_i as the solution to $z_i = \int_{s_{i-1}}^{s_i} \lambda(t \mid H_t)dt$.
4. If $u_i > T$ then stop.
5. $i = i+1$.
6. Go to step 2.

Example 1 (cont'd). Simulation of a Retinal Neuron

Figure 16 shows an example of a simulated spike train for the retinal neuron under constant light conditions. The spikes were generated by time-rescaling using the model of Eq. 12 and the parameters in Fig. 8. The conditional intensity function at each point in time was calculated based on the spiking history in the past 120 msec. The history was

initialized as 120 msec with no spikes, and each spike that was subsequently generated was added to this history. Visually, we can see that this simulated spike train has similar features to the real data shown in figure 2C. There are a number of bursts with multiple spikes being fired rapidly, as well as some longer ISIs.

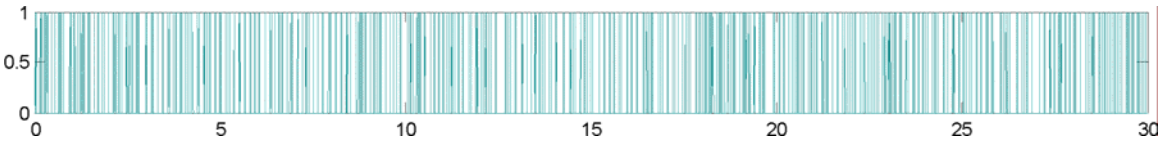

Figure 16. Simulated spontaneous spiking activity of a retinal neuron generated by time-rescaling.

Example 3 (cont'd). Simulation of a Hippocampal Neuron

Figure 17 shows an example of the simulated spiking activity of a hippocampal place cell with the same place field properties as the one shown in Figure 10B. The estimated parameters and the actual path of the rat were plugged into the model of equation 15 to obtain a firing rate at each point in time. The maximum firing rate was about 12.5 spikes/second. The spiking activity was simulated by thinning, using this maximum value. 10,000 spikes were generated, and of those, 1725 were accepted. It is clear from Figure 17 that even though the generated spikes were independent of the animal's position, these spike were much more likely to be accepted in the region of the place field.

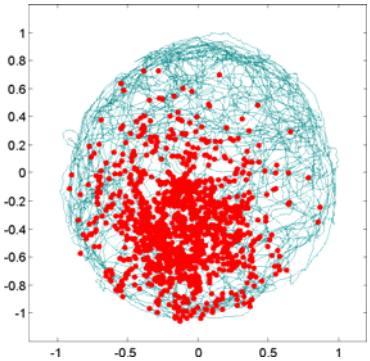

Figure 17. Simulated spiking activity of a place cell generated by thinning.

Appendix:

AI. The sum of independent Poisson random variables is Poisson

Let X and Y be independent Poisson random variables with parameters λ_x and λ_y respectively. Then the probability mass function for the random variable $Z = X + Y$ is

$$\begin{aligned}\Pr(Z=n) &= \sum_{k=0}^{n} \Pr(X=k)\Pr(Y=n-k) \\ &= \sum_{k=0}^{n} \frac{\lambda_x^k e^{-\lambda_x}}{k!}\frac{\lambda_y^{n-k} e^{-\lambda_y}}{(n-k)!} \\ &= \frac{e^{-(\lambda_x+\lambda_y)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_x^k\lambda_y^{n-k}.\end{aligned}$$

Applying the binomial theorem to the last expression gives,

$$\Pr(Z=n) = \frac{e^{-(\lambda_x+\lambda_y)}}{n!}(\lambda_x+\lambda_y)^n, \tag{A1}$$

which is the probability mass function for a Poisson random variable with parameter $\lambda_x+\lambda_y$. Therefore, the sum of two independent Poisson random variables is itself a Poisson random variable with parameter equal to the sum of the parameters of the original Poissons. It is clear that this result can be extended to the sum of any number of Poisson random variables, which will still be Poisson with a parameter that is the sum of all the component parameters.

This result is important to ensure that our first definition of a homogeneous Poisson process, which states that the distribution of any interval has a Poisson distribution, is well defined. For example, let $N(t)$ be the counting process for a simple Poisson process with rate λ. Then by this definition, $\Delta N_{(t,t+\Delta t]} \sim \text{Pois}[\lambda\Delta t]$ and $\Delta N_{(t+\Delta t,t+\Delta t+\Delta s]} \sim \text{Pois}[\lambda\Delta s]$. It must also be the case, by this definition, that $\Delta N_{(t,t+\Delta t+\Delta s]} \sim \text{Pois}[\lambda(\Delta t+\Delta s)]$. We see that this is indeed the case since $\Delta N_{(t,t+\Delta t+\Delta s]} = \Delta N_{(t,t+\Delta t]} + \Delta N_{(t+\Delta t,t+\Delta t+\Delta s]}$, which is the sum of Poisson random variables and is hence Poisson with parameter $\lambda(\Delta t+\Delta s)$. Therefore, it is necessarily true that if the increments over all smaller intervals are Poisson then so too are the increments over all other intervals.

Equation A1 is also important for understanding the properties of inhomogeneous Poisson processes. We originally defined this process as the limit of a Bernoulli process as $\Delta t \to 0$. Below, in section AII, we show that in this limit the Bernoulli increments are equivalent to Poisson increments. If we split any interval $(a,b]$ into increasingly finer standard partitions by letting n grow to infinity and by setting $\Delta t = (b-a)/n$, then $\Delta N_{(a,b]} = \lim_{n\to\infty}\sum_{i=1}^{n}\Delta N_{(a+(i-1)\Delta t,a+i\Delta t]}$, which is the sum of infinitesimal Poisson increments. By Equation A1, this must have a Poisson distribution with a parameter equal to $\lim_{n\to\infty}\sum_{i=1}^{n}\lambda(a+i\Delta t)\Delta t$. But this limit is by definition the Riemann integral of $\lambda(t)$ over the interval $(a,b]$. Therefore, $\Delta N_{(a,b]} \sim \text{Pois}[\int_a^b \lambda(t)dt]$.

AII. As Δt approaches 0, Poisson and Binomial increments become equivalent.

Let $o(\Delta t)$ denote any function such that $\lim_{\Delta t \to 0} o(\Delta t) / \Delta t = 0$. In other words, a function is $o(\Delta t)$ if it goes to zero faster than Δt itself. We show that the difference between the probability mass function of a Poisson and a Bernoulli random variable is $o(\Delta t)$, and this component becomes negligible as $\Delta t \to 0$.

Assume ΔN_i is a Poisson increment with parameter $\lambda \Delta t$. Then the probability of observing no spikes in the interval is $\Pr(\Delta N_i = 0) = e^{-\lambda \Delta t} = 1 - \lambda \Delta t + o(\Delta t)$, where the final term comes from the Taylor expansion of $e^{-\lambda \Delta t}$. Up to order $o(\Delta t)$, this is equivalent to the probability of a Bernoulli$[p]$ random variable being 0, when $p = \lambda \Delta t$. The probability of firing exactly one spike in the interval is $\Pr(\Delta N_i = 1) = \lambda \Delta t e^{-\lambda \Delta t} = \lambda \Delta t (1 - \lambda \Delta t + o(\Delta t)) = \lambda \Delta t + o(\Delta t)$, which is equivalent to the probability of a Bernoulli$[p]$ random variable being 1, when $p = \lambda \Delta t$. And the probability of firing more than one spike is $\Pr(\Delta N_i = k) = (\lambda \Delta t)^k e^{-\lambda \Delta t} / k! = o(\Delta t)$, for $k > 1$, which becomes negligibly small as Δt approaches 0.

Therefore, for small enough Δt, we can treat a sequence of Poisson increments as a simple Bernoulli process with binary outcomes in each interval. In particular, this fact means that Definitions 1 and 2 for a simple Poisson process in Section IV are equivalent. For inhomogeneous Poisson processes, this means that for small enough intervals, each increment is Poisson, and hence by the discussion in Section AI, the increments over longer intervals are also Poisson.

AIII. The ISI distribution and joint density for a general point process can be computed from the Bernoulli approximation

For any interval, $(a, b]$, we construct the n^{th} standard partition of that interval by setting $\Delta t = (b - a)/n$, and letting $t_i = a + i \Delta t$. The probability of observing zero spikes in the larger interval is equivalent to the probability of observing no spikes in each of the partition intervals,

$$\begin{aligned} \Pr(\Delta N_{(a,b]} = 0) &= \Pr(\Delta N_{(t_0,t_1]} = 0, \ldots, \Delta N_{(t_{n-1},t_n]} = 0) \\ &= \Pr(\Delta N_{(t_{n-1},t_n]} = 0 \mid H_{n-1}) \cdots \Pr(\Delta N_{(t_0,t_1]} = 0 \mid H_0). \end{aligned}$$

Equation 5 states that for sufficiently small Δt, each of these small history dependent increments takes on the value 0 with probability $1 - \lambda(t_i \mid H_i)\Delta t$. Therefore,

$$\begin{aligned}\Pr(\Delta N_{(a,b]}=0) &= \lim_{\Delta t\to 0}\prod_k (1-\lambda(t_k \mid H_k)\Delta t) \\ &= \lim_{\Delta t\to 0}\prod_k (\exp(-\lambda(t_k \mid H_k)\Delta t)+o(\Delta t)) \\ &= \lim_{\Delta t\to 0}\exp(-\sum_k \lambda(t_k \mid H_k)\Delta t)+o(\Delta t) \\ &= \exp(-\int_a^b \lambda(t \mid H)dt),\end{aligned} \tag{A2}$$

where the limit of the sum in the exponential term is the Riemann integral of the conditional intensity function over $(a,b]$.

Assume that we observe the $(i-1)^{\text{th}}$ spike of a spike train at a time s_{i-1}, and the preceding spike history, $H_{s_{i-1}}$, and we want to compute the probability distribution of the next spike time, S_i. The probability that S_i is greater than some time s is just the probability that the point process does not spike between s_{i-1} and s. Therefore by Equation A2,

$$\Pr(S_i > s \mid H_{s_{i-1}}) = \exp\left(-\int_{s_{i-1}}^{s} \lambda(t \mid H_t)dt\right),$$

and the CDF of the i^{th} spike time is

$$\Pr(S_i \le s \mid H_{s_{i-1}}) = 1-\exp\left(-\int_{s_{i-1}}^{s} \lambda(t \mid H_t)dt\right).$$

We obtain the pdf of the i^{th} spike time by differentiating the CDF with respect to s,

$$f_{S_i}(s \mid H_{s_{i-1}}) = \frac{d}{ds}\left(1-\exp\left(-\int_{s_{i-1}}^{s} \lambda(t \mid H_t)dt\right)\right) = \lambda(s \mid H_s)\exp\left(-\int_{s_{i-1}}^{s} \lambda(t \mid H_t)dt\right). \tag{A3}$$

This is Equation 6 in section VI.

Using Equations A2 and A3, we can compute the joint likelihood of observing any spike train.

$$\begin{aligned}f_{S_1,\ldots,S_{N(T)}}\left(s_1,\ldots,s_{N(T)}\right) &= \prod_{i=1}^{N(T)}\left(f_{S_i}(s_i \mid H_{s_{i-1}})\right)\Pr(\Delta N_{(S_{N(T)},T]}=0) \\ &= \prod_{i=1}^{N(T)}\left(\lambda(s_i \mid H_{s_i})\exp\left\{-\int_{s_{i-1}}^{s_i}\lambda(t \mid H_t)dt\right\}\right)\exp\left\{-\int_{s_{N(T)}}^{T}\lambda(t \mid H_t)dt\right\} \\ &= \prod_{i=1}^{N(T)}\left(\lambda(s_i \mid H_{s_i})\right)\exp\left\{-\int_0^T \lambda(t \mid H_t)dt\right\}\end{aligned} \tag{A4}$$

This is Equation 7 in section VI. Once a conditional intensity model is specified for a point process, and data is observed, this joint density is the likelihood function for that

data, as a function of any parameters in the model. This likelihood will be the cornerstone of our point process model fitting framework.

AIV. Proof of the time rescaling theorem

Let $(0,T]$ be an observation interval and let $N(t)$ for $0 < t \le T$ be the counting process of a point process with conditional intensity function $\lambda(t \mid H_t)$. Assume we observe a set of spikes $0 < s_1 < ... < s_{N(T)} \le T$. Define z_i for $i = 1,..., N(T)$ as in Equation #####. Define a Jacobian matrix whose i, j^{th} element is

$$J_{i,j} = \frac{dz_i}{ds_j} = \begin{cases} \lambda(s_i \mid H_i) & \text{if } i = j \\ -\lambda(s_j \mid H_j) & \text{if } i = j-1 \\ 0 & \text{otherwise} \end{cases},$$

This Jacobian is lower triangular, so its determinant is just the product of its diagonal elements, $|J| = \prod_{i=1}^{N(T)} \lambda(s_i \mid H_{s_i})$.

Using Equation 6, the joint density function of the first n spike times is given by

$$\begin{aligned} f_{S_1,...,S_n}(s_1,...,s_n) &= \prod_{i=1}^{n} f_{S_i}(s_i \mid H_{s_{i-1}}) \\ &= \prod_{i=1}^{n} \left(\lambda(s_i \mid H_{s_i}) \exp\left\{ -\int_{s_{i-1}}^{s_i} \lambda(t \mid H_t) dt \right\} \right) \end{aligned}$$

According to the change of variables formula, the joint density function of the random variables $Z_i = \int_{S_{i-1}}^{S_i} \lambda(t \mid H_t) dt$, for $i = 1,..., N(T)$ is given by

$$\begin{aligned} f_{Z_1,...,Z_{N(T)}}(z_1,...,z_{N(T)}) &= f_{S_1,...,S_{N(T)}}(s_1,...,s_{N(T)}) |J|^{-1} \\ &= \prod_{i=1}^{N(T)} \left(\lambda(s_i \mid H_{s_i}) \exp\left\{ -\int_{s_{i-1}}^{s_i} \lambda(t \mid H_t) dt \right\} \right) \left(\prod_{i=1}^{N(T)} \lambda(s_i \mid H_{s_i}) \right)^{-1} \\ &= \prod_{i=1}^{N(T)} \left(\exp\{-z_i\} \right), \end{aligned}$$

which is the joint probability density of independent exponential random variables with unit rate parameter.