

Chapter 19

Point Processes

At the beginning of this book, in Example 1.1 (p. 3), we described the activity of a neuron recorded from the supplementary eye field. Interpreting Fig. 1.1 we said that, toward the end of each trial, the neuron fired more rapidly under one experimental condition than under the other. In that discussion we took for granted one of the foundational teachings¹ of neurophysiology, that neurons respond to a stimulus or contribute to an action by increasing their firing rate. But what, precisely, do we mean by “firing rate?” The definition of firing rate turns out to be both subtle and important for statistical analysis of neural data.

Perhaps the simplest conception is that firing rate (FR) is number of spikes (action potentials) per unit time. To compute it we would then count spikes over a time interval of length Δt and write

$$FR = \frac{\text{number of spikes}}{\Delta t}. \tag{19.1}$$

While useful in many contexts, Eq. (19.1) suffers from a fundamental difficulty: it depends strongly on the interval used in the calculation. As an extreme case, suppose we were to examine an interval of length $\Delta t = 100$ ms containing a single spike. Rewriting in terms of seconds, we get $\Delta t = .1$ s (seconds) and this would give us $FR = 10$ spikes per second (10 Hz). But now suppose we shrink the interval down to $\Delta t = 5$ ms. Then we would have $\Delta t = .005$ s and we would get $FR = 200$ Hz, which is drastically different. How would we know what interval to choose?

To avoid this conundrum, and to begin the process of formulating a statistical model, we do two things. First, we replace the spike count by its theoretical counterpart, the expected spike count, and then we pass to the limit as $\Delta t \rightarrow 0$ so that we obtain a firing rate at time t that no longer involves an interval. In other words, we define a theoretical *instantaneous firing rate*. Note that for small Δt the

¹ Description of this phenomenon began with work of Edgar Adrian and Keffer Hartline and their colleagues (e.g., Adrian and Zotterman 1926; Hartline and Graham 1932).

spike count in (19.1) is either 0 or 1, which is a Bernoulli event with expected value $P(\text{spike in } (t, t + \Delta t))$. The theoretical instantaneous firing rate at time t then becomes

$$FR(t) = \lim_{\Delta t \rightarrow 0} \frac{P(\text{spike in } (t, t + \Delta t))}{\Delta t}. \quad (19.2)$$

However, the definition in (19.2) omits any mention of the experimental context of the observed firing rate. A more inclusive way to write firing rate as a function of time is to allow it to depend on variables we write, collectively, as a vector x_t . The vector x_t might refer to an experimental condition or it could involve such things as refractory effects due to a previous spike shortly before time t (see Section 19.1.3), or a local field potential that represents a substantial component of synaptic input to the cell. We therefore have a more complete conceptualization of firing rate by putting it in the form

$$FR(t|x_t) = \lim_{\Delta t \rightarrow 0} \frac{P(\text{spike in } (t, t + \Delta t)|x_t)}{\Delta t}. \quad (19.3)$$

To flesh this out we must say how we calculate the probability in the numerator of (19.3), which will take us through Section 19.3.2. Granting that we will get there, we may state the central idea in statistical modeling of spike train data: neurophysiological phenomena may be represented through variables x_t that are thought to influence spiking activity. A statistical model for spike trains involves two things: (1) a simple, universal formula for the probability density of the spike train in terms of the instantaneous firing rate function, and (2) a specification of the way the firing rate function depends on variables x_t .

A major theme of this book is the use of probability to describe variation. In Chapter 3 we considered events, which led to our description of variation using probability distributions, and in Chapter 18 we examined sequences of temporally-dependent observations, which were modeled as time series. Spike trains, however, don't quite fit into any of the molds we have constructed in the foregoing chapters. They are sequences of varying *event times*, times at which action potentials (spikes) occur—in repeated trials the spike times typically vary, as may be seen in Fig. 1.1 of Example 1.1. To handle such sequences of event times we invoke a special class of models called *point processes*. As we discuss in Section 19.3.4, the tools needed for fitting point processes to spike train data are generalized linear models (Chapter 14) and nonparametric regression (Chapter 15). Indeed, the models we discuss that involve instantaneous firing rate, conceptualized by (19.3), are called *point process regression models*. The purposes of this chapter are, first, to review the way point process representations of spike trains are defined in terms of instantaneous firing-rate functions and, second, to show how point process regression models help in understanding neural behavior.

The name “point process” reflects the localization of the events as points in time together with the notion that the probability distributions evolve across time according to a *stochastic process*. Point processes can be more general, so that the points

can lie in a higher-dimensional physical or abstract space. In PET imaging, for example, a radioisotope that has been incorporated into a metabolically active molecule is introduced into the subject's bloodstream and after these molecules become concentrated in specific tissues the radioisotopes decay, emitting positrons which may be detected. These emissions represent a four-dimensional *spatiotemporal* point process because they are localized occurrences both spatially, throughout the tissue, and in time. Here, however, we focus on point processes in time and their application to modeling spike trains.

The simplest point processes are *Poisson processes*, which are *memoryless* in the sense that the probability of an event occurring at a particular time does not depend on the occurrence or timing of past events. In Section 19.2.1 we discuss *homogeneous* Poisson processes, which can describe highly irregular sequences of event times that have no discernible temporal structure. When an experimental stimulus or behavior is introduced, however, time-varying characteristics of the process become important. In Section 19.2.2 we discuss Poisson processes that are *inhomogeneous* across time. In Section 19.3 we describe ways that more general processes can retain some of the elegance of Poisson processes while gaining the ability to describe a rich variety of phenomena.

Spike trains are fundamental to information processing in the brain, and point processes form the statistical foundation for distinguishing signal from noise in spike trains. We have already seen in Chapters 14 and 15 examples of spike train analysis using Poisson regression with spike counts. For this purpose, the Poisson regression model may be conceptualized as involving counts observed over time bins of width Δt based on a neural firing rate $FR(t)$. In Poisson regression, each Poisson distribution has mean equal to $FR(t) \cdot \Delta t$ and then $FR(t)$ is related to the stimulus (or the behavior) by a formula we may write in short-hand as

$$\log FR(t) = \text{stimulus effects}, \quad (19.4)$$

meaning that $\log FR(t)$ is some function that is determined by the stimulus or behavior. In Example 14.5, for instance, the right-hand side of (19.4) involved a quadratic function that represented the effective distance of a rat from the preferred location of a particular hippocampal place cell, and the result was a Poisson regression model of the place cell's activity. This sort of model may be considered a kind of simplified prototype. When we pass to the limit as in (19.2) and use instantaneous firing rate, the Poisson regression model becomes a Poisson process regression model.

Poisson processes are important, and they are especially useful for analyzing the trial-averaged firing rate. When, in Example 15.1, we displayed the smoothed PSTH under two experimental conditions, we were comparing two trial-averaged firing-rate functions. We spell this out in Section 19.3.3. On the other hand, many phenomena can only be studied *within trials*. For instance, oscillatory behavior, bursting, and some kinds of influences of one neuron on another show substantial variation across trials and may be difficult or impossible to detect from across-trial summaries like the PSTH. Careful examination of spike trains within trials usually reveals non-Poisson behavior: neurons tend not to be memoryless, but instead exhibit effects

of their past *history* of spiking (e.g., of refractory effects or recent burst activity). Non-Poisson models that incorporate history effects are described in Section 19.3, and methods developed in that section produce within-trial analyses of spike trains. In such cases, the instantaneous firing rate takes the form (19.3) and Eq. (19.4) must be modified by including additional terms (as components of the variable x_t) on the right-hand side to incorporate effects that occur differently on each trial. For instance, a firing-rate model might have the form

$$\log FR(t|x_t) = \text{stimulus effects} + \text{history effects} + \text{coupling effects}. \quad (19.5)$$

In Section 19.3.4 we indicate how spike train data may be analyzed by fitting models suggested by conceptualizations like (19.5), again using the methods developed in Chapters 14 and 15.

19.1 Point Process Representations

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

If s_1, s_2, \dots, s_n are times at which events occur within some time interval we may take $x_i = s_i - s_{i-1}$, i.e., x_i is the elapsed time between s_{i-1} and s_i , and define $x_1 = s_1$. This gives the inter-event waiting times x_i from the event times and we could reverse the arithmetic to find the event times from a set of inter-event waiting times x_1, \dots, x_n using $s_j = \sum_{i=1}^j x_i$. In discussing point processes, both of these representations are useful. In the context of spike trains, s_1, s_2, \dots, s_n are the spike times, while x_1, \dots, x_n are the inter-spike intervals (ISIs). Nearly all of our discussion of event-time sequences will involve modeling of spike train behavior.

To represent the variability among the event times we let X_1, X_2, \dots be a sequence of positive random variables. Then the sequence of random variables S_1, S_2, \dots defined by $S_j = \sum_{i=1}^j X_i$ is a *point process* on $(0, \infty)$. In fitting point processes to data, we instead consider finite intervals of time over which the process is observed, and these are usually taken to have the form $(0, T]$, but for many theoretical purposes it is more convenient to assume the point process ranges across $(0, \infty)$.

Another useful way to describe a set of event times is in terms of the counts of events observed over time intervals. The event count in a particular time interval may be considered a random variable. For theoretical purposes it is helpful to introduce a function $N(t)$ that counts the total number of events that have occurred up to and including time t . $N(t)$ is called the *counting process* representation of the point process. See Fig. 19.1. If we let $\Delta N_{(t_1, t_2]}$ denote the number of events observed in the interval $(t_1, t_2]$, then we have $\Delta N_{(t_1, t_2]} = N(t_2) - N(t_1)$. The count $\Delta N_{(t_1, t_2]}$ is often called the *increment* of the point process between t_1 and t_2 . In the case of a

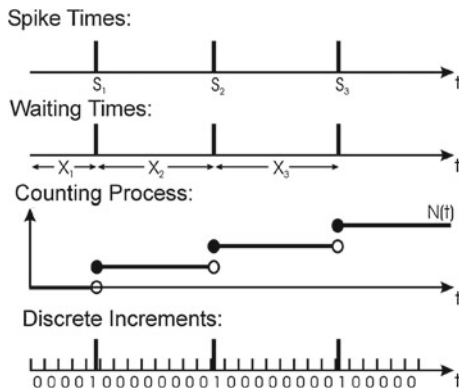


Fig. 19.1 Multiple specifications for point process data: the process may be specified in terms of spike times, waiting times, counts, or discrete binary indicators.

neural spike train, S_i would represent the time of the i th spike, X_i would represent the i th inter-spike interval (ISI), and $\Delta N_{(t_1, t_2]}$ would represent the spike count in the interval $(t_1, t_2]$. For event times S_i and inter-event waiting times X_i we are dealing with mathematical objects that are already familiar, namely sequences of random variables, with the index i being a positive integer. The counting process, $N(t)$, on the other hand, is a *continuous-time stochastic process*, which determines count increments that are random variables.

Keeping track of the times at which the count increases is equivalent to keeping track of increments. Furthermore, for successive spike times s_i and s_{i+1} , if we set $t_1 = s_i$ and consider $t_2 < s_{i+1}$ then $\Delta N_{(t_1, t_2]} = 0$ but when $t_2 = s_{i+1}$ then $\Delta N_{(t_1, t_2]} = 1$. Thus, keeping track of the times at which the count increases is equivalent to keeping track of events themselves and, therefore, the counts provide a third way to characterize a point process.

As an example of the way we may identify the event times with the counting process, the set of times for which the counting process is less than some value j , $\{t : N(t) < j\}$, is equivalent to the set of times for which the j th spike has not yet occurred, $\{t : S_j > t\}$. Both of these representations express the set of all times that precede the j th spike, but they do so differently. We can describe a point process using spike times, interspike intervals, or counting processes and specifying any one of these fully specifies the other two. It is often possible to simplify theoretical calculations by taking advantage of these multiple equivalent representations.

19.1.2 A point process may be considered, approximately, to be a binary time series.

At the beginning of the chapter we said that point process data are analyzed using the framework of generalized linear models. This requires the discrete representation

given at the bottom of Fig. 19.1. The event times, inter-event intervals, and counting process all specify the point process in *continuous* time. Suppose we take an observation interval $(0, T]$ and break it up into n small, evenly-spaced time bins. Let $\Delta t = T/n$, and $t_i = i \cdot \Delta t$, for $i = 1, \dots, n$. We can now consider the discrete increments $\Delta N_i = N(t_i) - N(t_{i-1})$, which count the number of events in a single bin. If we make Δt small enough, it becomes extremely unlikely for there to be more than one event in a single bin. The set of increments $\{\Delta N_i; i = 1, \dots, n\}$ then becomes a sequence of 0s and 1s, with the 1s indicating the bins in which the events are observed (see Fig. 19.1). In the case of spike trains, data are often recorded in this form, with $\Delta t = 1$ ms. To emphasize the point, we define $Y_i = \Delta N_i$, and put $p_i = P(Y_i = 1)$, so that $Y_i \sim \text{Bernoulli}(p_i)$. The Y_i s form a binary time series, that is, a sequence of Bernoulli random variables that may be inhomogeneous (the p_i may be different) and/or dependent. Such a discrete-time process is yet another way to represent a point process, at least approximately. It loses some information about the precise timing of events within each bin, but for sufficiently small Δt this loss of information becomes irrelevant for practical purposes. Also, for small Δt we have small p_i and the Bernoulli distributions may be approximated by Poisson distributions, according to the result in Section 5.2.2. In other words, for small Δt we may consider the point process to be essentially a sequence of Poisson random variables. This will allow us to use Poisson regression methods (which are part of generalized linear model methodology) in analyzing data modeled as point processes. The rest of this chapter is largely devoted to filling in the details and fleshing out the consequences, thereby supplying the substance behind the informal statements (19.4) and (19.5).

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

In many stochastic systems, past behavior influences the future. The biophysical properties of ion channels, for example, make it impossible for a neuron to fire again immediately following a spike, creating a short interval known as the absolute refractory period. In addition, after the absolute refractory period there is a relative refractory period during which the neuron can fire again, but requires stronger input in order to do so. These refractory effects are important cases of *history dependence* in neural spike trains. To describe spike train variability accurately (at least for moderate to high firing rates where the refractory period is important), the probability of a spike occurring at a given time must depend on how recently the neuron has fired in the past. A more complicated history-dependent neural behavior is bursting, which is characterized by short sequences of spikes with small interspike intervals. In addition, spike trains are sometimes oscillatory. For example, neurons in the CA1 region of rodent hippocampus tend to fire at particular phases of the EEG theta rhythm. Thus,

in a variety of settings, probability models for spike trains make dependence on spiking history explicit.

Example 19.1 Retinal ganglion cell under constant conditions Neurons in the retina typically respond to patterns of light displayed over small sections of the visual field. When retinal neurons are grown in culture and held under constant light and environmental conditions, however, 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. A short segment of the spiking activity from one neuron appeared in Fig. 16.1. A histogram of the ISIs appears in the left panel of Fig. 19.10. Even though this neuron is not responding to any explicit stimuli, we can still see structure in its firing activity. Although most of the ISIs are shorter than 20 ms, some are much longer: there is a small second mode in the histogram around 60–120 ms. This suggests that the neuron may experience two distinct states, one in which there are bursts of spikes (with short ISIs) and another, more quiescent state (with longer ISIs). From Fig. 16.1 we may also get an impression that there may be bursts of activity, with multiple spikes arriving in quick succession of one another. □

Example 19.2 Beta oscillations in Parkinson's disease Parkinson's disease, a chronic progressive neurological disorder, causes motor deficits leading to difficulty in movement. Clinical studies have shown that providing explicit visual cues, as guides, can improve movement in many patients, a possible explanation being that cortical drive associated with cues may lead to dampening of pathological beta oscillations (10–30 Hz) in the basal ganglia. To investigate this phenomenon, Sarma et al. (2012) recorded from neurons in the basal ganglia (specifically, the substantia nigra) while patients carried out a hand movement task. Because the period associated with a 20 Hz oscillation is 50 ms, if a neuron's activity is related to a beta oscillation it will tend to fire roughly every 50 ms. Therefore, its probability of firing at time t will be elevated if it fired previously 50 ms prior to time t . This is a form of history effect, which the authors built into their neural models in order to examine whether it was dampened due to visual cues. □

Example 19.3 Spatiotemporal correlations in visual signaling To better understand the role of correlation among retinal ganglion cells, Pillow et al. (2008) examined 27 simultaneously-recorded neurons from an isolated monkey retina during stimulation by binary white noise. The authors used a model having the form of (19.5). They concluded, first, that spike times appear more precise when the spiking behavior of coupled neighboring neurons is taken into account and, second, that in predicting (decoding) the stimulus from the spike trains, inclusion of the coupling term improved prediction by 20% compared with a method that ignored coupling and instead assumed independence among the neurons. □

19.2 Poisson Processes

19.2.1 Poisson processes are point processes for which event probabilities do not depend on occurrence or timing of past events.

The discussion in Section 19.1.3 indicated the importance of history dependence in spike trains. On the other hand, a great simplification is achieved by ignoring history dependence and, instead, assuming the probability of spiking at a given time has no relationship with previous spiking behavior. This assumption leads to the class of *Poisson processes*, which are very appealing from a mathematical point of view: although they rarely furnish realistic models for data from individual spike trains, they are a pedagogical—and often practical—starting point for point processes in much the way that the normal distribution is for continuous random variables. As we shall see below, it is not hard to modify Poisson process models to make them more realistic.

Two kinds of Poisson processes must be distinguished. When event probabilities are invariant in time Poisson processes are called *homogeneous*; otherwise they are called *inhomogeneous*. We begin with the homogeneous case.

Definition: A homogeneous Poisson process with intensity λ is a point process satisfying the following conditions:

1. For any interval, $(t, t + \Delta t]$, $\Delta N_{(t, t + \Delta t]} \sim P(\mu)$ with $\mu = \lambda \Delta t$.
2. For any non-overlapping intervals, $(t_1, t_2]$ and $(t_3, t_4]$, $\Delta N_{(t_1, t_2]}$ and $\Delta N_{(t_3, t_4]}$ are independent.

For spike trains, the first condition states that for any time interval of length Δt , the spike count is a Poisson random variable with mean $\mu = \lambda \cdot \Delta t$. In particular, the mean, which is the expected number of spikes in the interval, increases in proportion to the length of the interval. Furthermore, the distribution of the spike count depends on the length of the interval, but not on its starting time: $\Delta N_{(t, t+h]}$ has the same distribution as $\Delta N_{(s, s+h]}$ for all positive values of s, t, h . This homogeneous process is *time-invariant*, and is said to have *stationary increments*. The second condition states that the spike counts (the counting process increments) from non-overlapping intervals are independent. In other words, the distribution of the number of spikes in an interval does not depend on the spiking activity outside that interval. Another way to state this definition is to say that a homogeneous Poisson process is a point process with stationary, independent increments.

A detail: There is one technical point to check: we need to be sure that the distributions of overlapping intervals, given in the definition

above, are consistent. For example, if we consider intervals (t_1, t_2) and (t_2, t_3) we must be sure that the Poisson distributions for the counts in each of these are consistent with the Poisson distribution for the count in the interval (t_1, t_3) . Specifically, in this case, we must know that the sum of two independent Poisson random variables with means $\mu = \lambda(t_2 - t_1)$ and $\mu = \lambda(t_3 - t_2)$ is a Poisson random variable with mean $\mu = \lambda(t_3 - t_1)$. But this follows from the fact that if $W_1 \sim P(\mu_1)$ and $W_2 \sim P(\mu_2)$ independently, and we let $W = W_1 + W_2$, then $W \sim P(\mu_1 + \mu_2)$. We omit the details. \square

We now come to an important characterization of homogeneous Poisson processes.

Theorem: A point process is a homogeneous Poisson process with intensity λ if and only if its inter-event waiting times are i.i.d. $Exp(\lambda)$.

Proof: We derive the waiting-time distribution for a homogeneous Poisson process. Recalling that X_i is the length of the inter-event interval between the $(i - 1)$ st and i th event times, we have $X_i > t$ precisely when $\Delta N_{(S_{i-1}, S_{i-1}+t]} = 0$. From the definition of a homogeneous Poisson process, $P(\Delta N_{(S_{i-1}, S_{i-1}+t]} = 0) = e^{-\lambda t}$. Therefore, the CDF of X_i is $F_{X_i}(t) = P(X_i \leq t) = 1 - e^{-\lambda t}$, which is the CDF of an $Exp(\lambda)$ random variable.

The converse of this theorem involves additional calculations and is omitted. \square

Recall from Section 5.4.2 that the exponential distribution is memoryless. According to this theorem, for a homogeneous Poisson process, at any given moment the time at which the next event will occur does not depend on past events. Thus, the homogeneous Poisson process “has no memory” of past events.

Another way to think about homogeneous Poisson processes is that the event times are scattered “as irregularly as possible.” One characterization of the “irregularity” notion is that, as noted on p. 120, the exponential distribution $Exp(\lambda)$ maximizes the entropy among all distributions on $(0, \infty)$ having mean $\mu = 1/\lambda$. Here is another.

Result: Suppose we observe $N(T) = n$ events from a homogeneous Poisson process on an interval $(0, T]$. Then the distribution of the event times is the same as that of a sample of size n from a uniform distribution on $(0, T]$.

Proof: This appears as a corollary to the theorem on p. 577, where it is also stated more precisely. \square

Example 19.4 Miniature excitatory post-synaptic currents Figure 19.2 displays event times of miniature excitatory postsynaptic currents (MEPSCs) recorded from neurons in neonatal mice at multiple days of development. To record these events, the neurons are patch clamped at the cell body and treated so that they cannot propagate action potentials. These MEPSCs are thought to represent random activations of the dendritic arbors of the neuron at distinct spatial locations, so that the two assumptions of a Poisson process are plausible. The sequence of events in Fig. 19.2 looks highly



Fig. 19.2 A sequence of MEPSC event times. The inter-event intervals are highly irregular.

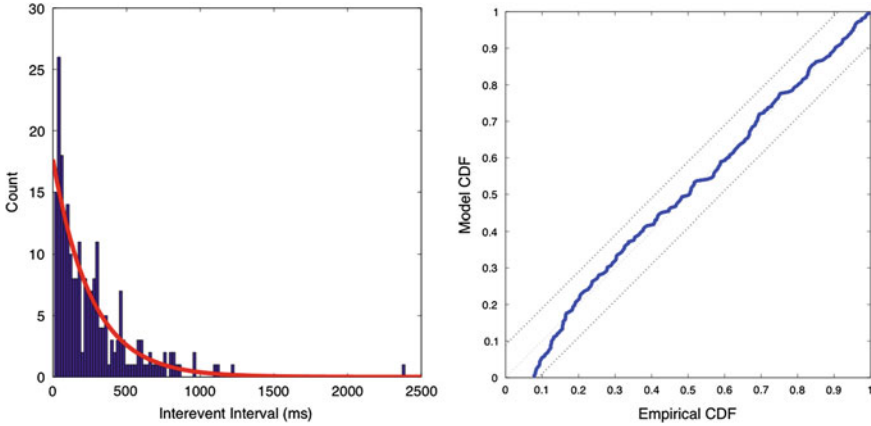


Fig. 19.3 Histogram and P–P plot of MEPSC inter-event intervals. *Left* Overlaid (in red) on the histogram is an exponential pdf. *Right* P–P plot falls within diagonal bands, indicating no lack of fit according to the Kolmogorov-Smirnov test (discussed in Section 19.3.5).

irregular, with no temporal structure. Figure 19.3 displays a histogram of the intervals between MEPSC events. The distribution of waiting times is captured well by an exponential fit, as shown both in left panel of Fig. 19.3 and in the P–P plot, in the right panel, which compares² the empirical CDF to that of an exponential. □

Important intuition may be gained by considering a discrete time representation of a sequence of event times, as discussed in Section 19.1.2. Suppose we have an observation interval $(0, T]$ and we consider partitioning $(0, T]$ into successive time bins of width Δt . If we make Δt sufficiently small we can force to nearly zero the probability of getting more than 1 event in any time bin. We then ignore the possibility of getting more than 1 event in any bin and, as in Section 19.1.2, we then let Y_i be the binary random variable that indicates whether an event has occurred in the i th time bin with $P(Y_i = 1) = p_i$, for $i = 1, \dots, n$ (so that there are n time bins and $T = n\Delta t$). Each Y_i is a *Bernoulli* (p_i) random variable. If these Bernoulli random variables are homogeneous ($p_1 = p_2 = \dots = p_n = p$ for some p) and independent, so that they form Bernoulli trials, then we have

1. For the i th time bin $(i\Delta t, (i + 1)\Delta t]$, $\Delta N_{(i\Delta t, (i+1)\Delta t)} \sim \text{Bernoulli}(p)$.

² The small deviation of the curve from the diagonal in the lower left-hand corner of the P–P plot is probably due to inaccuracy of measurement for very short inter-event intervals.

2. For any two distinct time bins, $(i\Delta t, (i+1)\Delta t]$ and $(j\Delta t, (j+1)\Delta t]$, $\Delta N_{(i\Delta t, (i+1)\Delta t)}$ and $\Delta N_{(j\Delta t, (j+1)\Delta t)}$ are independent.

Let us now put $\lambda = p/\Delta t$ and use the Poisson approximation to the binomial distribution (see Section 5.2.2) as $\Delta t \rightarrow 0$. The two properties above then become essentially (for sufficiently small Δt) the same as the two properties in the definition of a Poisson process, given on p. 570. Therefore, leaving aside some mathematical details (see (19.11)), we may say that the sequence of Bernoulli trials converges to a Poisson process as $\Delta t \rightarrow 0$. That is, a homogeneous Poisson process is essentially a sequence of Bernoulli trials. We used this idea repeatedly in interpreting the Poisson distribution in Section 5.2. Rewriting $\mu = p/\Delta t$ as $p = \lambda\Delta t$ and replacing Δt with the infinitesimal dt we obtain the shorthand summary

$$P(\text{event in } (t, t + dt)) = \lambda dt. \quad (19.6)$$

We extend the fundamental connection between Bernoulli random variables and Poisson processes (and therefore also Poisson distributions) to the inhomogeneous case in Section 19.2.2.

19.2.2 Inhomogeneous Poisson processes have time-varying intensities.

We made two assumptions in defining a simple Poisson process: that the increments were (i) stationary, and (ii) independent for non-overlapping intervals. The first step in modeling a larger class of point processes is to eliminate the stationarity assumption. For spike trains, we would like to construct a class of models where the spike count distributions vary across time. In terms of the Bernoulli-trial approximation, we wish to allow the event probabilities p_i to differ.

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

1. For any interval, $(t, t + \Delta t]$, $\Delta N_{(t, t + \Delta t]} \sim P(\mu)$ with $\mu = \int_t^{t+\Delta t} \lambda(t) dt$.
2. For any non-overlapping intervals, $(t_1, t_2]$ and $(t_3, t_4]$, $\Delta N_{(t_1, t_2]}$ and $\Delta N_{(t_3, t_4]}$ are independent.

This process is called an inhomogeneous Poisson process because 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. The inhomogeneous Poisson process is no longer stationary, but its increments remain independent and, as a result, it retains the memoryless property, according to which the probability of spiking at any instant

does not depend on occurrences or timing of past spikes. In shorthand notation we modify (19.6) by writing

$$P(\text{event in } (t, t + dt]) = \lambda(t)dt. \quad (19.7)$$

At the beginning of the chapter we said that point process data are analyzed using the framework of generalized linear models, and in Section 19.1.2 we identified as a key step the representation of a point process as a binary time series, at least approximately. To take this step we need to equate, at least approximately, the point process likelihood function and the likelihood function for a suitable binary time series. In general, a likelihood function is proportional to the joint pdf of the data. Suppose we have observed event times s_1, \dots, s_n . We assume these arise as observed values of random variables $S_1, \dots, S_{N(T)}$, where $N(T)$ is the number of event times in $(0, T]$ and is itself a random variable. We write the joint pdf of s_1, \dots, s_n as $f_{S_1, \dots, S_{N(T)}}(s_1, \dots, s_n)$, where we acknowledge in our subscript notation³ that $N(T)$ is also a random variable (taking the value $N(T) = n$ in data consisting of n events). Now suppose this joint pdf depends on some parameter vector θ . The likelihood function becomes

$$L(\theta) = f_{S_1, \dots, S_{N(T)}}(s_1, \dots, s_n | \theta). \quad (19.8)$$

In Example 14.5, for instance, we could consider the spike times to follow an inhomogeneous Poisson process and the parameter vector in (19.8) would consist of the parameters characterizing the spatial place cell distribution, $\theta = (\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy})$. To get a formula for the likelihood function, the mathematical result we need is the formula for the joint pdf of the spike times. To be sure we get essentially the same likelihood function when we instead treat the spike train as a binary time series we also need a statement that the joint pdf of the spike times is approximately equal to the joint pdf for the binary time series. We provide both of these results below. We then also present an additional fact about inhomogeneous Poisson processes that aids intuition.

We begin with the joint pdf.

Theorem The event time sequence $S_1, S_2, \dots, S_{N(T)}$ from a Poisson process with intensity function $\lambda(t)$ on an interval $(0, T]$ has joint pdf

$$f_{S_1, \dots, S_{N(T)}}(s_1, \dots, s_n) = \exp \left\{ - \int_0^T \lambda(t) dt \right\} \prod_{i=1}^n \lambda(s_i). \quad (19.9)$$

Proof: See Section 19.4. □

³ A more explicit notation would be $f_{S_1, \dots, S_{N(T)}, N(T)}(S_1 = s_1, \dots, S_{N(T)} = s_n, N(T) = n)$, see p. 577, where we make explicit the randomness due to $N(T)$.

We now turn to our ability to treat an inhomogeneous Poisson process as if it were approximately the same as a binary time series described in Section 19.1.2, with

$$P(\text{event in } (t, t + \Delta t]) \approx \lambda(t)\Delta t. \quad (19.10)$$

We give a rigorous statement that the joint pdf of the spike times is approximately equal to the joint pdf for the corresponding binary time series. More specifically, we show that the joint pdf in Eq. (19.9) is the limit of relevant binary pdfs as $\Delta t \rightarrow 0$.

Let us consider a set of points s_1, \dots, s_n in the interval $(0, T]$ that, while conceptually representing event times, are for the purposes of the analysis below, taken to be fixed. They represent the observed data. We will call them “atoms” because they are points where probability mass will be placed. Suppose $(0, T]$ is decomposed into N subintervals of length Δt , so that $\Delta t = T/N$. For $i = 1, \dots, N$ let $x_i = 1$ if the i th subinterval contains one of the atoms and 0 otherwise.

Theorem Let $\lambda(t)$ be a continuous function on $[0, T]$, set $\lambda_i = \lambda(t_i)$ for subinterval midpoints t_i , and let $p_i = (\Delta t)\lambda_i$. Then as $\Delta t \rightarrow 0$ we have

$$\frac{1}{(\Delta t)^n} \prod_{i=1}^n p_i^{x_i} (1 - p_i)^{1-x_i} \rightarrow e^{-\int_0^T \lambda(t) dt} \prod_{i=1}^n \lambda(s_i). \quad (19.11)$$

To prove this result we need two lemmas. Let $S = S_n$ be the set of i indices for which $x_i = 1$ and S^c the set of indices for which $x_i = 0$.

Lemma 1 As $\Delta t \rightarrow 0$ we have

$$\prod_S \lambda(t_i) \rightarrow \prod_{i=1}^n \lambda(s_i).$$

Proof: The lemma follows immediately from continuity of $\lambda(t)$. □

Lemma 2 As $\Delta t \rightarrow 0$ we have

$$\sum_{S^c} \log(1 - (\Delta t)\lambda_i) \rightarrow -\int_0^T \lambda(t) dt.$$

Proof: This follows immediately from a first-order Taylor series expansion of the log (Equation (A.5)), together with the definition of the integral as a limit⁴ of sums. □

Proof of the theorem: Putting the two lemmas together we easily prove the theorem. We have

⁴ The limit of the sum over S^c is the same as the limit of the sum over $S \cup S^c$ because S has n elements for all sufficiently small values of Δt , so that $\lim \sum_S \Delta t \lambda_i = 0$.

$$\begin{aligned}
\frac{1}{(\Delta t)^n} \prod_{i=1}^N p_i^{x_i} (1-p_i)^{1-x_i} &= \frac{1}{(\Delta t)^n} \left(\prod_S (\Delta t) \lambda_i \right) \left(\prod_{S^c} 1 - (\Delta t) \lambda_i \right) \\
&= \left(\prod_S \lambda_i \right) e^{\sum_{S^c} \log(1 - (\Delta t) \lambda_i)} \\
&\rightarrow e^{-\int_0^T \lambda(t) dt} \prod_{i=1}^n \lambda(s_i). \quad \square
\end{aligned}$$

To recap: taken together, the two theorems above show that the inhomogeneous Poisson process spike time joint pdf is approximately equal to a binary time series joint pdf, which allows us to use the binary random variables Y_i (with $p_i = P(Y_i = 1)$) defined in Section 19.1.2 in place of the Poisson process. The memorylessness of the Poisson process translates into independence among the Y_i s. However, the values of p_i may vary across time, corresponding to the inhomogeneity of the process. Importantly, we may estimate $\lambda(t)$ by likelihood methods, applying Poisson regression with suitably small time bins (e.g., having width 1 ms).

Example 1.1 (continued) In Chapter 1 we introduced the SEF neuron example, the problem being to characterize the neural response under two different experimental conditions. In Chapter 8 we returned to the example to describe the benefit of smoothing the PSTH, and in Chapter 15, p. 422, we showed how smoothing may be accomplished using Poisson regression splines. The smoothing model was

$$Y_i \sim P(\lambda_i) \quad (19.12)$$

$$\log \lambda_i = f(t_i) \quad (19.13)$$

where t_i was the time at the midpoint of the i th time bin (of the PSTH), Y_i was the corresponding spike count in that bin, and $f(t)$ was taken to be a natural cubic spline with two knots at specified locations.

An inhomogeneous Poisson process model may be constructed that is very similar to the PSTH-based regression model. To get a Poisson process model we must take the time bins to be smaller—small enough that on any trial there is at most one spike in any bin. For instance, we may take the bins to have width 1 ms. Then, we must define the resulting binary counts: for trial r let Y_{ri} be 1 if a spike occurs in the i th bin and 0 otherwise. We write the model

$$Y_{ri} \sim P(\lambda_i) \quad (19.14)$$

$$\log \lambda_i = f(t_i) \quad (19.15)$$

where, again, $f(t)$ is a natural cubic spline with two knots at the locations specified previously. Comparing (19.14) and (19.15) with (19.12) and (19.13) we have a model of almost the same form. Aside from the width of the time bins, the distinction is that (19.14) and (19.15) is a within-trial model, in terms of Y_{ri} , while (19.12) and (19.13) is a model that pools events across trials by using the PSTH spike counts Y_i . It turns

out that the intensity that results from fitting (19.14) and (19.15) is nearly identical to the fit of $f(t)$ resulting from (19.12) and (19.13). The closeness of results holds quite generally because the smoothing of the PSTH is not very sensitive to the choice of bin widths as long as the firing rate varies slowly enough to be nearly constant within bins. Smoothing the PSTH amounts to fitting a Poisson process after jittering all the spike times within a bin so that they are equal to the midpoint of that bin. \square

The final theorem of this section gives another interesting way to think about inhomogeneous Poisson processes. Let us begin by considering the PSTH, as used in Examples 1.1 and 15.1. The PSTH is the peristimulus time *histogram*. But in what sense is it a histogram? A histogram is a plot that displays counts, as does the PSTH, but the counts are presumed to be repeated observations from a random variable, and the histogram is supposed to be a rough estimate of the random variable's pdf. What are the repeated observations that generate the PSTH? And what pdf is it estimating? The data are the event times. But, as we have already taken pains to point out, these event times are not i.i.d. observations from a fixed distribution: they follow a point process, which is different. How are they transformed into i.i.d. observations that are suitable for making a histogram and estimating a pdf? While these questions are puzzling at first, the answer turns out to be simple. According to the next theorem, given some number n of events in an interval $(0, T]$, the event times will be scattered across $(0, T]$ as if they were i.i.d. observations from a distribution having as its pdf the normalized intensity $\lambda(t)$. In other words, the positions of the event times are just like i.i.d. observations; therefore, the PSTH is just like a histogram, and could be treated as if it were an estimator of the normalized intensity function.

To state the result, let us first recall that the length of the sequence of event times $S_1, S_2, \dots, S_{N(T)}$ depends on the random quantity $N(T)$. Thus, to be more thorough we might write the joint pdf above in the form

$$f_{S_1, \dots, S_{N(T)}}(s_1, \dots, s_n) = f_{S_1, \dots, S_{N(T)}, N(T)}(S_1 = s_1, \dots, S_{N(T)} = s_n, N(T) = n).$$

That is, the pdf on the left-hand side is really a short-hand notation for the pdf on the right-hand side. This observation is used in the proof of the following theorem. We will write $f_N(n)$ for the pdf of $N(T)$ and note that, for a Poisson process with intensity $\lambda(t)$, $N(T) \sim P(\mu)$ with $\mu = \int_0^T \lambda(t) dt$.

Theorem Let $S_1, S_2, \dots, S_{N(T)}$ be an event sequence from a Poisson process with intensity function $\lambda(t)$ on an interval $(0, T]$. Conditionally on $N(T) = n$, the sequence S_1, S_2, \dots, S_n , has the same joint distribution as an ordered set of i.i.d. observations from a univariate distribution having pdf

$$g(t) = \frac{\lambda(t)}{\int_0^T \lambda(u) du}.$$

Proof: We write the conditional pdf as

$$\begin{aligned}
f_{S_1, \dots, S_{N(T)}}(s_1, \dots, s_n | N(T) = n) &= \frac{f_{S_1, \dots, S_{N(T)}}(s_1, \dots, s_n)}{f_N(n)} \\
&= \frac{e^{-\int_0^T \lambda(t) dt} \prod_{i=1}^n \lambda(s_i)}{e^{-\int_0^T \lambda(t) dt} \frac{\left(\int_0^T \lambda(t) dt\right)^n}{n!}} \\
&= n! \prod_{i=1}^n \frac{\lambda(s_i)}{\int_0^T \lambda(t) dt} \\
&= n! \prod_{i=1}^n g(s_i).
\end{aligned}$$

Noting that there are $n!$ ways to order the observations s_1, \dots, s_n , this completes the proof. \square

The theorem says that we may consider an inhomogeneous Poisson process with intensity $\lambda(t)$ to be equivalent to a two-stage process in which we (1) generate an observation $N = n$ from a Poisson distribution with mean $\mu = \int_0^T \lambda(t) dt$; this tells us how many events are in $(0, T]$; we then (2) generate n i.i.d. observations from a distribution having $g(t) = \lambda(t) / \int_0^T \lambda(u) du$ as its pdf. We motivated the theorem by suggesting that it shows how the PSTH acts like a histogram: the intensity function $\lambda(t)$ describes the event times that come from pooling together all the spike times across all of the trials; the PSTH then estimates $\lambda(t) / \int_0^T \lambda(u) du$. Not only does this explain the sense in which the PSTH is actually a histogram, it also motivates application of a density estimator (e.g., a normal kernel density estimator or Gaussian filter), as in Section 15.4, to smooth the PSTH.

When we specialize the theorem above to homogeneous Poisson processes we get, as a corollary, the result stated as a theorem on p. 571.

Corollary Let $S_1, S_2, \dots, S_{N(T)}$ be an event sequence from a homogeneous Poisson process with intensity λ on an interval $(0, T]$. Conditionally on $N(T) = n$, the sequence S_1, S_2, \dots, S_n , has the same joint distribution as an ordered set of i.i.d. observations from a uniform distribution on $[0, T]$.

Proof: This is a special case of the theorem in which $\lambda(t) = \lambda$ so that $g(t) = 1/T$, i.e., $g(t)$ is the pdf of the uniform distribution on $(0, T]$. \square

19.3 Non-Poisson Point Processes

19.3.1 Renewal processes have i.i.d. inter-event waiting times.

The homogeneous Poisson process developed in Section 19.2.1 assumed that the point process increments were both stationary and independent of past event history.

To accommodate event probabilities that change across time, we generalized from homogeneous to inhomogeneous Poisson processes. This eliminated the assumption of stationary increments but it preserved the independence assumption, which entailed history independence. Systems that produce point process data, however, typically have physical mechanisms that lead to history-dependent variation among the events, which cannot be explained with Poisson models. Therefore, it is necessary to further generalize by removing the independence assumption.

The simplest kind of history-dependent behavior occurs when the probability of the i th event depends on the occurrence time of the previous event s_{i-1} , but not on any events prior to that. If the i th waiting time X_i is no longer memoryless, then $P(X_i > t + h | X_i > t)$ may not be equal to $P(X_i > u + h | X_i > u)$ when $u \neq t$, but X_i is independent of event times prior to S_{i-1} , and is therefore independent of all waiting times X_j for $j < i$. Thus, the waiting time random variables are all mutually independent. In the time-homogeneous case, they also all have the same distribution. A point process with i.i.d waiting times is called a *renewal process*. We already saw that homogeneous Poisson processes have i.i.d. exponential waiting times. Therefore, renewal processes may be considered generalizations of homogeneous Poisson processes.

A renewal model is specified by the distribution of the inter-event waiting times. 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 positive values, but 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, which generalizes the exponential, may be used when one wants to describe interspike interval distributions using two parameters: the gamma shape parameter gives it flexibility to capture a number of characteristics that are often observed in point process data. 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 are actually Poisson. If the shape parameter is less than one, then the density drops off faster than an exponential. This can provide a rough description of ISIs when a neuron fires 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 describe the ISIs for a relatively regular spike train, such as those from a neuron having oscillatory input. Thus, this very simple class of distributions with only two parameters is capable of capturing, at least roughly, some interesting types of history dependent structure.

While the gamma distribution is simple and flexible, it doesn't have any direct connection with the physiology of neurons. For neural spiking data, a renewal model with a stronger theoretical foundation is the inverse Gaussian. As described in Section 5.4.6, the inverse Gaussian also has two parameters and is motivated by

the integrate-and-fire conception of neural spiking behavior. Thus, a renewal process with inverse Gaussian ISIs would be a simple yet natural model for neural activity in a steady state.

One way to quantify the regularity of a renewal process is through the ISI coefficient of variation. We noted in (3.14) that exponentially-distributed random variables have $CV = 1$, so this corresponds to a Poisson process. When $CV < 1$ the process is more regular than Poisson (as would be a spike train from an oscillatory neuron), while when $CV > 1$ the process is more irregular than Poisson (as would be a spike train from a bursty neuron). This regularity or irregularity of a renewal process will also be apparent in the distribution of counts and is often measured by the Fano factor,

$$F(t, t + \Delta t) = \frac{V(\Delta N_{(t, t + \Delta t)})}{E(\Delta N_{(t, t + \Delta t)})}.$$

For a Poisson process we have $F(t, t + \Delta t) = 1$. The counts will be relatively less dispersed for regular renewal processes, so that $F(t, t + \Delta t) < 1$, and more dispersed for irregular processes, so that $F(t, t + \Delta t) > 1$.

A general result that has implications for spike train analysis is the *renewal theorem*, which⁵ examines the expected number of events in an interval $(t, t + h]$ as $t \rightarrow \infty$. For a Poisson process with intensity λ we have $E(\Delta N_{(t, t + h)}) = \lambda h$, and the waiting time distribution is exponential with mean $\mu = 1/\lambda$. In other words, the expected number of events in $(t, t + h]$ is $\lambda h = h/\mu$, so that the expected number of events is just the length of the interval divided by the average waiting time for an event. For a renewal process the same statement is approximately true for large t .

Renewal Theorem Suppose a renewal process has waiting times with a continuous pdf and a mean μ . Defining $\lambda = 1/\mu$ we have

$$\lim_{t \rightarrow \infty} E(\Delta N_{(t, t + h)}) = \lambda h.$$

Proof: Omitted. □

Notice that if we take h sufficiently small in the renewal theorem, the count $\Delta N_{(t, t + h]}$ will, with high probability, be either 0 or 1 and then its expectation is $E(\Delta N_{(t, t + h)}) = P(\Delta N_{(t, t + h)} = 1)$. Thus, if we pick a large t and ask for the probability of an event in the infinitesimal interval $(t, t + dt]$ by ignoring the time of the most recent event and instead letting the renewal process start at time 0 and run until we get to time t , we find that (19.6) continues hold.

A related result arises when we consider what happens when we combine multiple renewal processes by pooling together all their event times. This sort of pooling occurs, for example, in a PSTH when multiple spike trains are collected across multiple trials: in making the PSTH every spike time is used but the trial on which it occurred is ignored. Such combination of point processes is called *superposition*. Specifically, if we have counting processes $N^i(t)$, for $i = 1, \dots, n$ then

⁵ A more general version of this result is often called *Blackwell's Theorem*.

$N(t) = \sum_{i=1}^n N^i(t)$ is the process resulting from superposition. First, we consider the Poisson case.

Theorem For $i = 1, \dots, n$, let $N^i(t)$ be the counting process representation of a homogeneous Poisson process having intensity λ_i . Then the point process specified by $N(t) = \sum_{i=1}^n N^i(t)$ is a homogeneous Poisson process having intensity $\lambda = \sum_{i=1}^n \lambda_i$.

Sketch of Proof: Because the sum of independent Poisson random variables is Poisson, condition 1 of the definition of a homogenous Poisson process is satisfied for the superposition process. Because condition 2 is satisfied for all n independent processes, it is also satisfied for the superposition process. \square

Result The superposition of a large number of independent renewal processes having waiting times with continuous pdfs and finite means is, approximately, a Poisson process.

Proof: The mathematics involved in stating this result precisely are rather intricate. We omit the proof, but offer the following heuristics to make the result plausible.

Suppose that the n independent renewal processes have mean waiting times $\mu_i = 1/\lambda_i$, for $i = 1, \dots, n$. Let us consider intervals $(t, t+h]$, with h so small that, with large probability, across all n processes at most 1 event occurs. Then the superposition increments $\Delta N_{(t,t+h]}$ are essentially binary variables. For the superposition to be Poisson, these binary variables must be homogeneous and independent. By the renewal theorem, for large t ,

$$P(\Delta N_{(t,t+h]}^i = 1) \approx \lambda_i h,$$

where $\lambda_i = 1/\mu_i$ and

$$P(\Delta N_{(t,t+h]}^i = 0) \approx 1 - \lambda_i h.$$

When we pool all the processes together, the event $\Delta N_{(t,t+h]} = 1$ will occur if at least one process has an event, and otherwise $\Delta N_{(t,t+h]} = 0$, which has probability

$$P(\Delta N_{(t,t+h]}=0) \approx (1 - \lambda_1 h)(1 - \lambda_2 h) \cdots (1 - \lambda_n h) \approx e^{-\lambda t} \approx 1 - \lambda h$$

and this, in turn, shows that

$$P(\Delta N_{(t,t+h]} = 1) \approx \lambda h,$$

as for a Poisson process, so that homogeneity holds, approximately. As far as independence is concerned, the key point is that the renewal processes are independent of one another, so that the only dependence in the superposition is due to events from the same process, which are very rare among the large numbers of events in the superposition process. That is, if we assume n is so large that, for all k , $P(\Delta N_{(t,t+h]} = 1) \gg P(\Delta N_{(t,t+h]}^k = 1)$, then when we consider two non-overlapping intervals $(t_1, t_1 + h]$ and $(t_2, t_2 + h]$, relative to the superposition process, the probability that the k th process has events in both intervals is negligible. This is another way of saying that the identity of events in the superposition gets washed out as the number of processes increases. \square

By combining this superposition result and the renewal theorem we obtain a practical implication: the superposition of multiple renewal processes will be approximately a Poisson process, but we can expect the approximation to be better for large t , after initial conditions die out. If, for example, we take multiple spike trains, and if time $t = 0$ has a physiological meaning related to the conditions of the experiment, then we may expect the initial conditions to affect the spike trains in a reproducible way from trial to trial so that even after pooling we might see non-Poisson behavior near the beginning of the trial; as such effects dissipate across time we would expect the pooled spike trains to exhibit Poisson-process-like variation.

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

In Section 19.2.2 we described the structure of an inhomogeneous Poisson process in terms of an intensity function that characterized the instantaneous probability of firing a spike at each instant in time, as in (19.6). In an analogous way, a general point process may be characterized by its *conditional intensity function*. Poisson processes are memoryless but, in general, if we want to find the probability of an event in a time interval $(t, t + \Delta t]$ we must consider the timing of the events preceding time t . Let us denote the number of events prior to t by $N(t-)$,

$$N(t-) = \max_{u < t} N(u).$$

We call the sequence of event times prior to time t the *history* up to time t and write it as $H_t = (S_1, S_2, \dots, S_{N(t-)})$. For a set of observed data we would write $H_t = (s_1, s_2, \dots, s_n)$ with the understanding that $N(t-) = n$. The conditional intensity function is then given by

$$\lambda(t|H_t) = \lim_{\Delta t \rightarrow 0} \frac{P(\Delta N_{(t,t+\Delta t]} = 1|H_t)}{\Delta t}, \quad (19.16)$$

where $P(\Delta N_{(t,t+\Delta t]} = 1|H_t)$ is the conditional probability of an event in $(t, t + \Delta t]$ given the history H_t . Taking Δt to be small we may rewrite Eq. (19.16) in the form

$$P(\Delta N_{(t,t+\Delta t]} = 1|H_t) \approx \lambda(t|H_t)\Delta t. \tag{19.17}$$

Or, in shorthand,

$$P(\text{event in } (t, t + dt]|H_t) = \lambda(t|H_t)dt, \tag{19.18}$$

which generalizes (19.6). According to (19.18) the conditional intensity function expresses the instantaneous probability of an event. It serves as the fundamental building block for constructing the probability distributions needed for general point processes.⁶ A mathematical assumption needed for theoretical constructions is that the point process is *orderly*, which means that for a sufficiently small interval, the probability of more than one event occurring is negligible. Mathematically, this is stated as

$$\lim_{\Delta t \rightarrow 0} \frac{P(\Delta N_{(t,t+\Delta t]} > 1|H_t)}{\Delta t} = 0. \tag{19.19}$$

This assumption is biophysically plausible for a point process model of a neuron because neurons have an absolute refractory period. In most situations, the probability of a neuron firing more than one spike is negligibly small for $\Delta t < 1$ ms.

Once we specify the conditional intensity for a point process, it is not hard to write down the pdf for the sequence of event times in an observation interval $(0, T]$. In fact, the argument is essentially the same as in the case of the inhomogeneous Poisson process, with the conditional intensity $\lambda(t|H_t)$ substituted for the intensity $\lambda(t)$. The key observation is that the conditional intensity behaves essentially like a hazard function, the only distinction being the appearance of the stochastic history H_t .

Theorem The event time sequence $S_1, S_2, \dots, S_{N(T)}$ of an orderly point process on an interval $(0, T]$ has joint pdf

$$f_{S_1, \dots, S_{N(T)}}(s_1, \dots, s_n) = \exp \left\{ - \int_0^T \lambda(t|H_t)dt \right\} \prod_{i=1}^n \lambda(s_i|H_{s_i}) \tag{19.20}$$

where $\lambda(t|H_t)$ is the conditional intensity function of the process.

⁶ Because the history $H_t = (S_1, S_2, \dots, S_{N(t-)})$ is itself a point process, it is stochastic and, therefore, the conditional intensity is stochastic. The definition (19.18) includes two separable steps: first, we define the conditional intensity

$$\lambda(t|s_1, \dots, s_n) = \lim_{\Delta t \rightarrow 0} \frac{P(\Delta N_{(t,t+\Delta t]} = 1|N(t-) = n, S_1 = s_1, \dots, S_n = s_n)}{\Delta t}$$

for every possible vector (s_1, \dots, s_n) making up the history H_t , and then we replace the specific values $N(t-) = n$ and $(S_1 = s_1, \dots, S_n = s_n)$ with their stochastic counterparts written as $H_t = (S_1, S_2, \dots, S_{N(t-)})$.

Proof: See Section 19.4. □

Equation (19.20) has the same form as (19.9), the only distinction being the replacement of the Poisson intensity $\lambda(t)$ in (19.9) with the conditional intensity $\lambda(t|H_t)$ in (19.20).

We may also approximate a general point process by a binary process. For small Δt , the probability of an event in an interval $(t, t + \Delta t]$

$$P(\text{event in } (t, t + \Delta t]|H_t) \approx \lambda(t|H_t)\Delta t \quad (19.21)$$

and the probability of no event is

$$P(\text{no event in } (t, t + \Delta t]|H_t) \approx 1 - \lambda(t|H_t)\Delta t. \quad (19.22)$$

Equation (19.21) generalizes (19.10). If we consider the discrete approximation, analogous to the Poisson process case, we may define $p_i = \int \lambda(t|H_t)dt$ where the integral is over the i th time bin. We again get Bernoulli random variables Y_i with $P(Y_i = 1) = p_i$ but now these Y_i random variables are *dependent*, e.g., we may have $P(Y_i = 1|Y_{i-1} = 1) \neq p_i$. The theorem giving (19.11) holds again when we replace $\lambda(t)$ with $\lambda(t|H_t)$. In practice, spike train analyses using dependent binary variables are a little more complicated than those using independent binary variables, but it remains relatively easy to formulate history-dependent models for these dependent variables by following a regression strategy that is very similar to that used previously, on p. 576. We give examples in Section 19.3.4.

19.3.3 *The marginal intensity is the expectation of the conditional intensity.*

Equation (19.16) gave the definition of the conditional intensity function. We now define the unconditional or *marginal intensity function* as

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(\Delta N_{(t, t+\Delta t]} = 1)}{\Delta t}. \quad (19.23)$$

Definition (19.23) may be rewritten in some informative ways. First, note that if X is a binary random variable its expectation is $E(X) = P(X = 1)$, as in (15.2). For Δt sufficiently small, $\Delta N_{(t, t+\Delta t]}$ is a binary random variable so that (19.23) may be written

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{E(\Delta N_{(t, t+\Delta t]})}{\Delta t}. \quad (19.24)$$

That is, the marginal intensity is the expected spike count density.

Next, according to the law of total probability (p. 86), for a pair of random variables Y and X and an event A we have $P(X \in A) = E_Y(P(X \in A|Y))$. Letting H_t play the role of Y and $\Delta N_{(t,t+\Delta t]} = 1$ the role of $X \in A$, we get, similarly,

$$P(\Delta N_{(t,t+\Delta t]} = 1) = E_{H_t} (P(\Delta N_{(t,t+\Delta t]} = 1|H_t))$$

and

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{E_{H_t} (P(\Delta N_{(t,t+\Delta t]} = 1|H_t))}{\Delta t}.$$

By interchanging⁷ the expectation and limiting operation we may then write

$$\lambda(t) = E_{H_t}(\lambda(t|H_t)). \quad (19.25)$$

Equation (19.25) explains the name “marginal” intensity. The intensity $\lambda(t)$ is marginal in much the same sense as when we have a pair of random variables (X, Y) and speak of the distribution of X as a marginal distribution because it is derived by averaging over all possible values of Y . Here, $\lambda(t)$ results from averaging the conditional intensity over all possible histories H_t . In the case of spike trains, the conditional intensity would apply to individual trials, while the marginal intensity would be the theoretical time-varying firing rate after averaging across trials. Importantly, we may consider $\lambda(t)$ to be the function being estimated by the PSTH. This does not require us to assume the trials are in any sense all the same. There could be some source of trial-to-trial variation, or even systematic variation (such as effects associated with learning across trials). Consideration of $\lambda(t)$ takes place whenever the average across trials seems meaningful and interesting.

As in Eq. (19.17) we may also write

$$P(\Delta N_{(t,t+\Delta t]} = 1) \approx \lambda(t)\Delta t \quad (19.26)$$

and we have the shorthand

$$P(\text{event in } (t, t + dt]) = \lambda(t)dt, \quad (19.27)$$

keeping in mind that we also take the left-hand side to mean

$$P(\text{event in } (t, t + dt]) = E_{H_t}P(\text{event in } (t, t + dt]|H_t).$$

Equation (19.27) must be compared with (19.18) and, of course, it has the same form as (19.6). We may therefore think of the average across histories (for spike trains, the average across trials) as defining a theoretical inhomogeneous Poisson process intensity. This is the intensity that is estimated by the PSTH.

⁷ General theory justifying the interchange of limit and expectation applies here.

The distinction between conditional and marginal intensities is so important for spike train analysis that we emphasize it, as follows.

If we consider spike trains to be point processes, within trials the instantaneous firing rate is $\lambda(t|H_t)$ and we have

$$P(\text{spike in } (t, t + dt)|H_t) = \lambda(t|H_t)dt,$$

while the across-trial average firing rate is $\lambda(t)$ and we have

$$P(\text{spike in } (t, t + dt)) = \lambda(t)dt.$$

19.3.4 Conditional intensity functions may be fitted using Poisson regression.

On p. 576 we discussed the way Poisson regression may be used to fit inhomogeneous Poisson process models. The key theoretical result that made this possible was Eq. (19.11) in conjunction with (19.10). As we said on p. 584, that theorem holds again for conditional intensity functions using Eq. (19.21). This means that Poisson regression can again be used for non-Poisson point processes.

We now give some examples in which conditional intensity functions have been fitted to spike train data.

Example 19.1 (continued from p. 569) Let us take time bins to have width $\Delta t = 1$ ms and write $\lambda_k = \lambda(t_k|H_{t_k})$, where t_k is the midpoint of the k th time bin. Defining

$$\log \lambda_k = \alpha_0 + \sum_{j=1}^{120} \alpha_j \Delta N_{(k-j-1, k-j]}, \quad (19.28)$$

we get a model with 120 history-related explanatory variables, each indicating whether or not a spike was fired in a 1 ms interval at a different time lag. The parameter α_0 provides the log background firing rate in the absence of prior spiking activity within the past 121 ms. Using Poisson regression with ML estimation (as in Section 14.1) we obtained $\hat{\alpha}_0 = 3.8$ so that, if there were no spikes in the previous 121 ms, the conditional intensity would become $\lambda_k = \exp(\hat{\alpha}_0) = 45$ spikes per second, corresponding to an average ISI of 22 ms. The MLEs $\hat{\alpha}_i$ obtained from the data are plotted in Fig. 19.4, in the form $\exp\{\hat{\alpha}_i\}$. The $\hat{\alpha}_i$ values related to 0–2 ms after a spike are large negative numbers, so that $\exp\{\hat{\alpha}_i\}$ is close to zero, leading to a refractory period when the neuron is much less likely to fire immediately after

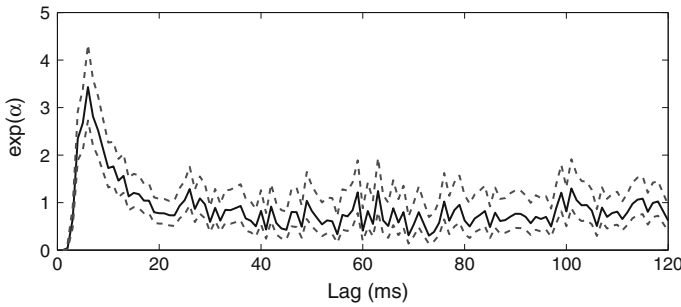


Fig. 19.4 Parameter estimates for history-dependent retinal conditional intensity model (*bold line*) together with confidence intervals (*dotted line*), which indicate uncertainty in the estimates (based on maximum likelihood). The x -axis indicates the lag time in milliseconds.

another spike. However, the estimates related to 4–13 ms after a spike are substantially positive, leading to an increase in the firing probability. For example, if the only spike in the 120 ms history occurred 6 ms 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. (The same data were discussed in the context of burst detection in Example 16.3 on p. 458.) Many of the remaining parameters are close to zero, and hence $\exp\{\hat{\alpha}_i\}$ is close to one, indicating that the corresponding history term has no effect on the spiking probability. Figure 19.5 displays the ISI histogram with exponential and Inverse Gaussian renewal model pdfs overlaid, and also the pdf for the model of Eq. (19.28). The exponential model overestimates the number of very short ISIs (0–4 ms), and both renewal models underestimate the number of ISIs between 5–10 ms and overestimate the number of ISIs between 10–60 ms. In contrast, the conditional intensity model in Eq. (19.28) accurately predicts the number of ISIs across all ISI lengths. □

Example 19.2 (continued) On p. 569 we said that a beta oscillation at 20 Hz could be represented in the history effects as an elevated probability of firing at time t when the neuron fired previously 50 ms prior to time t . Using Eq. (19.28) this would be represented by positive α_j coefficients around $j = 50$. Sarma et al. reduced the number of parameters, replacing (19.28) with

$$\log \lambda_k = \alpha_0 + \sum_{j=1}^{10} \alpha_j \Delta N_{k-j} + \sum_{i=1}^{14} \gamma_i \Delta N_{(k-(10i+9), k-10i]}. \tag{19.29}$$

In this version of the model, when γ_i is positive there is an increase in the log probability of firing when the neuron previously fired in the interval from $10i$ to $10i + 9$ ms in the past. Thus, the presence of a beta oscillation would produce a positive coefficient γ_5 (corresponding to 50–59 ms in the past, or 17–20 Hz). An example of a neuron having a positive γ_5 coefficient was given by the authors, reproduced here

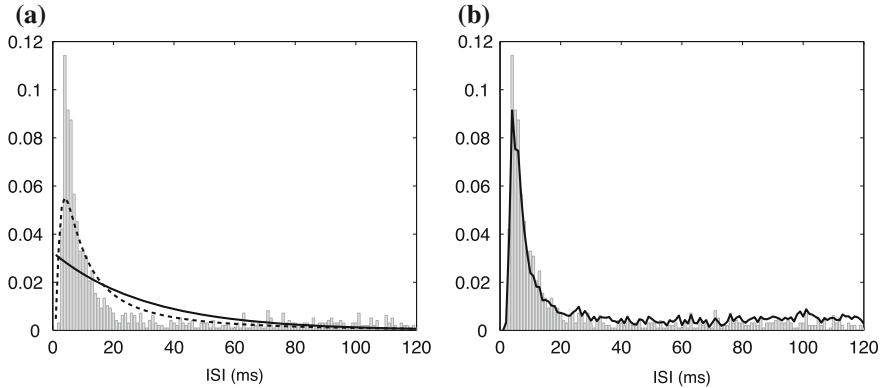


Fig. 19.5 ISI histogram and fitted pdfs. Panel **a**: ISI histogram overlaid with pdfs from exponential (solid line) and inverse Gaussian (dashed line) renewal models. Panel **b**: ISI histogram overlaid with pdf (solid line) from model defined by Eq. (19.28).

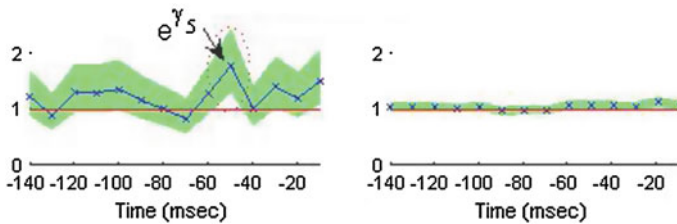


Fig. 19.6 Plots of γ coefficients using model (19.29) for a neuron recorded from the substantia nigra for a cued hand movement. *Left* coefficients before initiation of movement. *Right* coefficients after initiation of movement. Adapted from Sarma et al.

in Fig. 19.6. Results before and after movement initiation are shown in Fig. 19.6, when an explicit visual cue showed the subject where to move. In this case there was a dampening of beta oscillations during movement. The authors decomposed the timing of beta oscillations further and found that, among many substantia nigra cells, there was evidence of decreased beta oscillation beginning immediately following illumination of the visual cue. Based on additional results they suggested that execution of a motor plan following a cue may be suppressing pathological activity in the substantia nigra, which may explain improved task performance. \square

A second way to introduce history dependence is to begin with the hazard function of a renewal process and then modify the conditional intensity so that it can vary across time. This extends to renewal processes the method used for allowing Poisson processes to become inhomogeneous. In a homogeneous Poisson process, the waiting times are not only i.i.d., they are also memoryless: the probability of an event does not depend on when the last event occurred. To get an inhomogeneous Poisson process, we retain the memorylessness but introduce a time-varying conditional intensity. A simple idea is to take a renewal process and, similarly, introduce a time-varying

factor. For a renewal process, the probability of an event at time t depends on the timing of the most recent previous event $s_*(t)$, but not on any events prior to $s_*(t)$. If we allow the conditional intensity to depend on both time t and the time of the previous event $s_*(t)$ we obtain a form

$$\lambda(t|H_t) = g(t, s_*(t)) \quad (19.30)$$

where $g(x, y)$ is a function to be specified. Models of this type are sometimes called *Markovian* or *Inhomogeneous Markov Interval* (IMI) models.⁸ In an inhomogeneous Poisson process the conditional intensity takes the form

$$\lambda(t|H_t) = g_0(t)$$

where $g_0(t)$ becomes the intensity $\lambda(t)$. In a renewal process the conditional intensity takes the form

$$\lambda(t|H_t) = g_1(t - s_*(t))$$

where $g_1(t - s_*(t))$ becomes the hazard function for the waiting time distribution. The IMI model generalizes both of these, creating an inhomogeneous version of a renewal model.⁹ The simplest IMI model takes the conditional intensity to be of the multiplicative form¹⁰

$$\lambda(t|H_t) = g_0(t)g_1(t - s_*(t)). \quad (19.31)$$

A point process having conditional intensity of the form (19.30) or (19.31) may be fitted using binary Poisson regression, as in Example 1.1 on p. 576, except now with the additional terms representing the function $g_1(u)$ (where $u = t - s_*(t)$). A simple method is to fit the functions $g_0(t)$ and $g_1(u)$ using Poisson regression splines, in much the same way as discussed previously on p. 422 and 576 for Example 1.1.

Example 1.1 (continued from p. 576) Kass and Ventura (2001) fitted a model of the form (19.31) to data from an SEF neuron recorded for the study of Olson et al (2000). To do this they wrote

$$\log \lambda(t|H_t) = \log g_0(t) + \log g_1(t - s_*(t))$$

⁸ The terminology is intended to signify that the history dependence is limited to the previous spike time. A discrete-time stochastic process is a Markov process if the probability that the process will be in a particular state at time t depends only on the state of the process at time $t - 1$.

⁹ Because integrate-and-fire neurons reset to a baseline subthreshold voltage after firing, they necessarily follow Eq. (19.30). Further discussion of IMI models and their relationship to integrate-and-fire neurons is given in Koyama and Kass (2008).

¹⁰ The functions $g_0(t)$ and $g_1(u)$ are defined only up to a multiplicative constant. That is, for any nonzero number c if we multiply $g_0(t)$ by c and divide $g_1(u)$ by c we do not change the result. Some arbitrary choice of scaling must therefore be introduced. In Fig. 19.7 the constant was chosen so that $g_0(t)$ was equal to the Poisson process intensity at time $t = 50$ ms after the appearance of the visual cue.

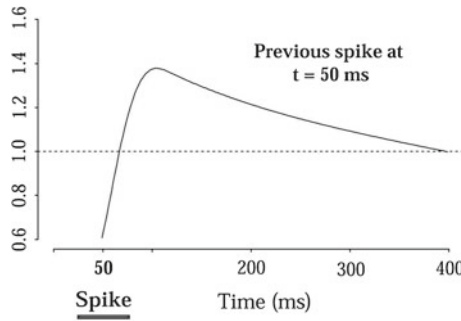


Fig. 19.7 Plot of the function $g_1(t - s_{\ast}(t))$ defined in (19.31) for the SEF data. The function is scaled so that a value of 1 makes the conditional intensity equal to the Poisson process intensity at time $t = 50$ ms after the appearance of the visual cue. Adapted from Kass and Ventura (2001).

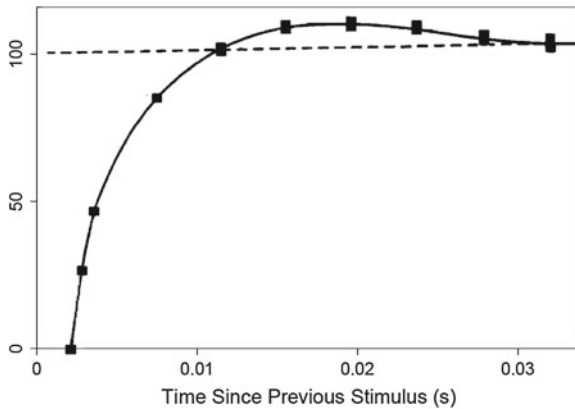


Fig. 19.8 Refractory effects in sciatic nerve of a frog. The y-axis is the reciprocal of the voltage threshold required to induce a second spike following a previous spike. The value 100 on the y-axis indicates the required reciprocal voltage when there was a long gap between the two successive action potentials. Adapted from Adrian and Lucas (1912).

which is an instance of (19.5) without coupling terms. Kass and Ventura took both $\log g_0(t)$ and $\log g_1(u)$ to be splines with a small number of knots and applied Poisson regression (see p. 422) using standard software. They showed that the model fitted the data better than an inhomogeneous Poisson model (using the graphical method in Section 19.3.5), and that inclusion of cross-product terms did not improve the fit (the likelihood ratio test for the additional terms was not significant).

A plot of the resulting non-Poisson recovery function $g_1(u)$ is shown in Fig. 19.7. For a Poisson process this function would be constant and equal to 1. The plot shows neural firing to be inhibited, compared with Poisson, for about 10 milliseconds and then it becomes *more* likely to fire, with the increase declining gradually until it returns to a baseline value. □

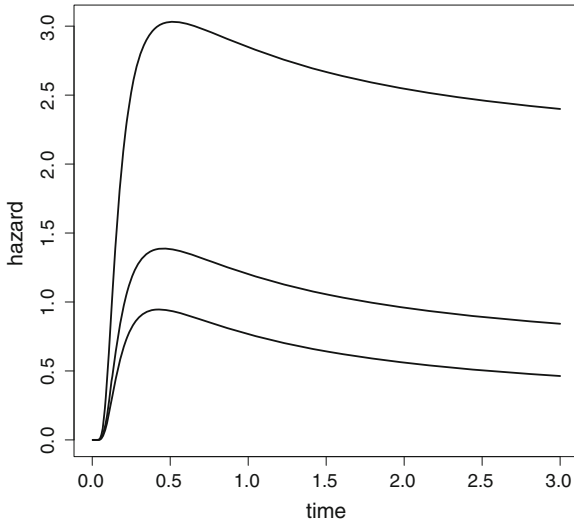


Fig. 19.9 Plots of inverse Gaussian hazard function for three different values of the coefficient of variation, .7 (*top curve*), 1 (*middle curve*), and 1.3 (*bottom curve*). These values correspond to the rough range of those commonly observed in cortical interspike interval data. The theoretical coefficient of variation is given by Eq. (5.16).

The non-monotonic behavior of the recovery function $g_1(t - s_*(t))$ in the foregoing analysis of Example 1.1 may seem somewhat surprising, but anecdotal evidence suggests it may be common. Interestingly, Adrian and Lucas (1912) found a qualitatively similar result by a very different method. They stimulated a frog's sciatic nerve through a second electrode and examined the time course of "excitability," which they defined as the reciprocal of the voltage threshold required to induce an action potential. Figure 19.8 plots this excitability as a function of time since the previous stimulus. There is again a relative refractory period of approximately 10 ms followed by an overshoot and a gradual return to the baseline value. Furthermore, the theoretical inter-spike interval distribution for an integrate-and-fire neuron (following a random walk generated by excitatory and inhibitory post-synaptic potentials) is inverse Gaussian (see Section 5.4.6), and the hazard function for an inverse Gaussian has a non-monotonic shape, shown in Fig. 19.9, that closely resembles the typical recovery function. The qualitative shape of the recovery function shown in Fig. 19.7 is thus consistent with what we would expect from the point of view of theoretical neurobiology.

In many experimental settings spike trains are collected to see how they differ under varying experimental conditions. The conditions may be summarized by a variable or vector, often called a *covariate* (as in regression, see p. 332). Furthermore, there may be other variables that may be related to spiking activity, which could be time-varying, such as a local field potential. Let us collect any such covariates into a vector denoted by u_t if we regard them as fixed by the experimenter, and V_t if

they should be considered stochastic. We then write $X_t = (H_t, u_t, V_t)$ and let the conditional intensity become a function not only of time and history, but also of the covariate vector X_t . Thus, for an observation $X_t = x_t$ we write the conditional intensity in the form $\lambda(t|x_t)$. With this in hand we may generalize the statement on p. 586, allowing it to cover the interesting cases implied by our discussion surrounding Eq. (19.5), as follows:

If we consider spike trains to be point processes, within trials the instantaneous firing rate is $\lambda(t|x_t)$ and we have

$$P(\text{spike in } (t, t + dt)|H_t) = \lambda(t|x_t)dt. \quad (19.32)$$

We may also generalize formula (19.20).

Theorem If the conditional intensity of an orderly point process on an interval $(0, T]$ depends on the random process X_t , so that when $X_t = x_t$ it may be written in the form $\lambda(t|x_t)$, then, conditionally on $X_t = x_t$, the event time sequence $S_1, S_2, \dots, S_{N(T)}$ has joint pdf

$$f_{S_1, \dots, S_{N(T)}|X_t}(s_1, \dots, s_n|X_t = x_t) = \exp\left\{-\int_0^T \lambda(t|x_t)dt\right\} \prod_{i=1}^n \lambda(s_i|x_t). \quad (19.33)$$

Proof: The proof is the same as that given for (19.20) in Section 19.4 with x_t replacing H_t . \square

A detail: If we are interested in the variation of the conditional intensity with the random vector X_t we can emphasize this by writing it in the form $\lambda(t|X_t)$. For example, in a multi-trial experiment, the firing rate may vary across trials, and the conditional intensity could include a component that changes across trials (see Ventura et al. 2005b). In such situations, the model includes two distinct sources of variability: one due to variability described by the point process pdf in (19.33) and the second due to the way the conditional intensity varies with X_t . The resulting point process is often called *doubly stochastic*. \square

Example 16.6 (continued from p. 472) We now give some additional details about the model used by Frank et al (2002). They applied a multiplicative IMI model to characterize spatial receptive fields of neurons from both the CA1 region of the hippocampus and the deep layers of the entorhinal cortex (EC) in awake, behaving rats. In their model, each neuronal spike train was described in terms of a conditional intensity function of the form (19.31), where the temporal factor $g_0(t)$ became

$$g_0(t) = g^S(t, u_t)$$

where u_t is the animal's two-dimensional spatial location at time t . In other words, $g^S(t, u_t)$ is a time-dependent place field. As we said on p. 472 the authors adopted a state-space model (see Section 16.2.4), where the state variables involved features of the place field. By modeling the resulting conditional intensity in the form

$$\lambda(t|x_t) = g^S(t, u_t)g_1(t - s_*(t))$$

the authors found consistent patterns of plasticity in both CA1 hippocampal neurons and deep entorhinal cortex (EC) neurons, which were distinct: the spatial intensity functions of CA1 neurons showed a consistent increase over time, whereas those of deep EC neurons tended to decrease. They also found that the ISI-modulating factor $g_1(t - s_*(t))$ of CA1 neurons increased only in the “theta” region (75–150 ms), whereas those of deep EC neurons decreased in the region between 20 and 75 ms. In addition, the minority of deep EC neurons whose spatial intensity functions increased in area over time fired in a more spatially specific manner than non-increasing deep EC neurons. This led them to suggest that this subset of deep EC neurons may receive more direct input from CA1 and may be part of a neural circuit that transmits information about the animal's location to the neocortex. \square

It is easy to supplement (19.31) with terms that consider not only the spike $s_*(t)$ immediately preceding time t , but also the spike $s_{2*}(t)$ preceding $s_*(t)$, $s_{3*}(t)$ preceding $s_{2*}(t)$, etc. One way to do this is to write

$$\lambda(t|H_t) = g_0(t)g_1(t - s_*(t))g_2(t - s_{2*}(t))g_3(t - s_{3*}(t)) \quad (19.34)$$

or, equivalently,

$$\begin{aligned} \log \lambda(t|H_t) &= \log g_0(t) + \log g_1(t - s_*(t)) \\ &\quad + \log g_2(t - s_{2*}(t)) + \log g_3(t - s_{3*}(t)) \end{aligned}$$

and then use additional spline-based terms to represent $\log g_2(t - s_{2*}(t))$ and $\log g_3(t - s_{3*}(t))$ in a Poisson regression.

Example 1.1 (continued) In their study of the model (19.31) for SEF neurons, described on p. 589, Kass and Ventura also used a model that included several spikes preceding time t , as in (19.34). The implementation again used splines with a small number of knots to represent each of the additional functions $g_2(t - s_{2*})$, $g_3(t - s_{3*})$, etc. The authors found the extra terms did not improve the fit (the likelihood ratio test was not significant). \square

A detail: In applying (19.34) using regression splines, Kass and Ventura allowed the functions $g_1(t - s_*)$, $g_2(t - s_{2*})$, $g_3(t - s_{3*})$, to be distinct. A plausible alternative is to assume they have the same functional form, which would mean that they have the same knots and the

same coefficients. This would say that the way a spike at time s prior to time t alters the probability of neural firing at time t depends only on $t - s$ and not on how many spikes occur between time s and time t . In this case (19.34) is replaced by

$$\lambda(t|H_t) = g_0(t)g_1(t - s_*(t))g_1(t - s_{2*}(t))g_1(t - s_{3*}(t)).$$

This simplification reduces the number of parameters in the model. Models of this type were used by Pillow et al (2008). \square

Another way model (19.31) may be extended is to include terms corresponding to coupling between neurons, as indicated by (19.5). To illustrate, we may consider the effect of neuron B on a given neuron A by letting $u_*(t)$ be the time of the neuron B spike that precedes time t and, similarly, letting $u_{2*}(t)$ and be the time of the spike preceding $u_*(t)$ and $u_{3*}(t)$ the time of the spike preceding $u_{2*}(t)$. Then we may append to (19.34) a series of factors that represent the coupling effects. In logarithmic form, considering 3 spikes back in time, this becomes

$$\begin{aligned} \log \lambda(t|H_t) &= \log g_0(t) + \log g_1(t - s_*(t)) \\ &\quad + \log g_2(t - s_{2*}(t)) + \log g_3(t - s_{3*}(t)) \\ &\quad + \log h_1(t - u_*(t)) + \log h_2(t - u_{2*}(t)) \\ &\quad + \log h_3(t - u_{3*}(t)). \end{aligned} \tag{19.35}$$

Once again (19.35) takes the form of (19.5), and some version of Poisson regression may be applied.

Example 19.3 (continued) In introducing this example on p. 569 we said that the authors used a model having the form of (19.5). Let us be somewhat more specific. In terms of (19.35), Pillow et al. took the receptive-field stimulus effects ($g_0(t)$, here spatio-temporal as in Example 16.6) to be linear, i.e., a linear combination of 5×5 stimulus pixel intensities across 30 time bins. For the history effects and the coupling effects they did not use splines but rather used an alternative set of basis functions such that $\log \lambda(t|H_t)$ remained linear, as it does with regression splines in (19.35). They then applied Poisson regression. However, because their model involved a large number of free parameters they had to use a modified fitting criterion (a form of penalized fitting similar to that used with smoothing splines) which is beyond the scope our presentation here. \square

19.3.5 Graphical checks for departures from a point process model may be obtained by time rescaling.

As described in Section 3.3.1, Q-Q and P-P plots may be used to check the fit of a probability distribution to data. These plots indicate the discrepancy between the

empirical cdf $\hat{F}(x)$ and the theoretical cdf $F(x)$, the idea being that when $\hat{F}(x)$ is based on i.i.d. random variables we have $\hat{F}(x) \rightarrow F(x)$ for all x (if the distribution is continuous) as the sample size grows indefinitely large. In the case of point processes we may examine the inter-event waiting times X_1, \dots, X_n . For a homogeneous Poisson process these are i.i.d. $Exp(\lambda)$. Thus, to assess the fit of a homogeneous Poisson process to a sequence of event times we may simply compute the inter-event waiting times and examine a Q-Q or P-P plot under the assumption that the true waiting-time distribution is exponential. For an inhomogeneous Poisson process, or a more general point process, the waiting times are no longer i.i.d. Thus, this method can not be applied in the same form. However, a version of the probability integral transform (p. 122) may be used to create a homogeneous Poisson process from *any* point process. We begin with a conditional intensity function in the general form of Eq. (19.32).

Time Rescaling Theorem. Suppose we have a point process with conditional intensity function $\lambda(t|x_t)$ on $(0, T]$ and with occurrence times $0 < S_1 < S_2, \dots, < S_{N(T)} \leq T$. Suppose further that the waiting time distributions are continuous with $f_{X_j|S_{j-1}}(x) > 0$ on $(s_{j-1}, T]$, for all $j \geq 1$. If we define

$$Z_1 = \int_0^{S_1} \lambda(t|x_t) dt \quad (19.36)$$

and

$$Z_j = \int_{S_{j-1}}^{S_j} \lambda(t|x_t) dt \quad (19.37)$$

for $j = 2, \dots, N(T)$, then $Z_1, \dots, Z_{N(T)}$ are i.i.d. $Exp(1)$ random variables.¹¹

Proof: See Section 19.4. □

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 $\lambda(t|x_t)$ were constant and equal to one everywhere, then the process would be a homogeneous Poisson process with independent, exponential ISIs, and time does not need to be rescaled. When $\lambda(t|x_t)$ is less than one, the transformed event times z_j accumulate slowly and represent a shrinking of time, so that distant event times are brought closer together. Likewise, when $\lambda(t|x_t)$ is greater than one, the event times z_j accumulate more rapidly and represent a stretching of time, so that neighboring event times are drawn further apart.

With time rescaling in hand, we may now apply Q-Q or P-P plots to detect departures from a point process model: using the conditional intensity function we transform the time axis and judge the extent to which the resulting waiting times deviate from those predicted by an $Exp(1)$ distribution. Furthermore, in conjunction with a P-P plot, the Kolmogorov-Smirnov test (Section 10.3.7) may be applied to test

¹¹ Extending the argument slightly to include the interval (s_N, T) it may also be shown that $Z_1, \dots, Z_{N(T)}$ follow a homogeneous Poisson process with intensity $\lambda = 1$.

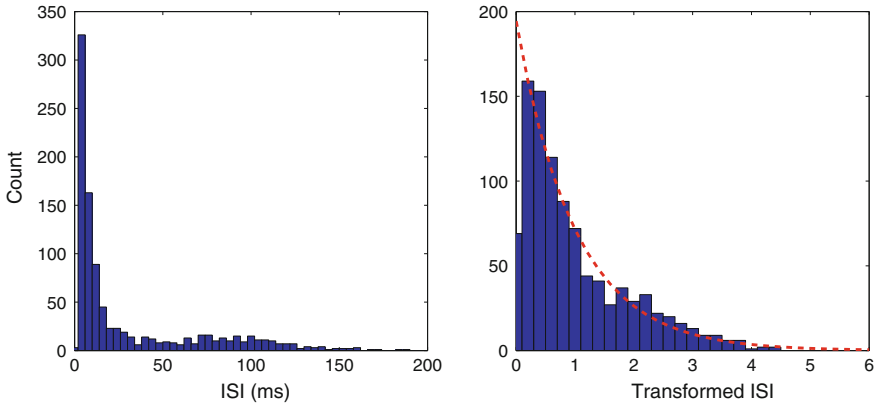


Fig. 19.10 *Left* Histogram of ISIs for the retinal ganglion cell spike train. *Right* Histogram of time-rescaled ISIs. *Dashed red line* is the $Exp(1)$ pdf.

the null hypothesis that the transformed waiting times follow an $Exp(1)$ distribution, which becomes an assessment of fit of the conditional intensity function. If the P–P plot consists of pairs (x_r, y_r) , for $r = 1, \dots, n$, the usual approach is to use the points $(x_r, y_r + 1.36/\sqrt{n})$ and $(x_r, y_r - 1.36/\sqrt{n})$ to define upper and lower bands for visual indication of fit, as illustrated in Fig. 19.11. Specifically, to make a P–P plot for a conditional intensity function $\lambda(t|x_t)$ used to model spike times s_1, s_2, \dots, s_n we do the following:

1. From (19.36) and (19.37) find transformed spike times z_1, \dots, z_n ;
2. for $j = 1, \dots, n$ define $u_j = 1 - \exp(-z_j)$;
3. put the values u_1, \dots, u_n in ascending order to get $u_{(1)}, \dots, u_{(n)}$;
4. for $r = 1, \dots, n$ (see p. 67) plot the (x, y) pair $\left(\frac{r-.5}{n}, u_{(r)}\right)$;
5. produce upper and lower bands: for $r = 1, \dots, n$ plot the (x, y) pair $\left(\frac{r-.5}{n}, u_{(r)} + 1.36/\sqrt{n}\right)$ and $\left(\frac{r-.5}{n}, u_{(r)} - 1.36/\sqrt{n}\right)$.

Example 19.1 (continued from p. 586) Using the conditional intensity of Eq. (19.28) we may apply time rescaling. Figure 19.10 displays 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 ms 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 the right panel of the figure shows the result of transforming the observed ISIs according to the conditional intensity model. Figure 19.11 displays a P–P plot for the intervals in the right panel of Fig. 19.10. Together, these figures show that the model in Eq. (19.28) does a good job of describing the variability in the retinal neuron spike train. \square

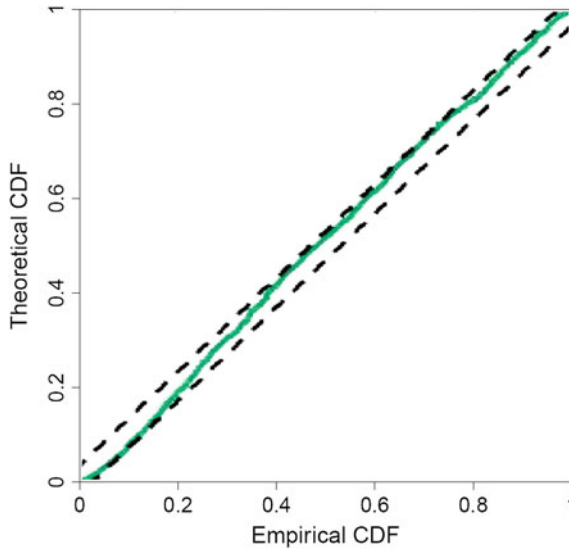


Fig. 19.11 P–P plot for the distribution of rescaled intervals shown in Fig. 19.10.

Example 19.5 Spike trains from a locust olfactory bulb. Substantial insight about sensory coding has been gained by studying olfaction among insects. An insect may come across thousands of alternative odors in its environment, among millions of potential possibilities, but only particular odors are important for the animal’s behavior. A challenge has been to describe the mechanisms by which salient odors are learned. A series of experiments carried out by Dr. Mark Stopfer and colleagues (e.g., Stopfer et al. 2003) has examined the way neural responses to odors may evolve over repeated exposure. To capture subtle changes it is desirable to have good point process models for olfactory spike trains. Figure 19.12 displays P–P plots for the fit of an inhomogeneous Poisson model and a multiplicative IMI model to a set of spike trains from a locust olfactory bulb. The spike trains clearly deviate from the Poisson model; the fit of the multiplicative IMI model to the data is much better. \square

19.3.6 There are efficient methods for generating point process pseudo-data.

It is easy to devise a computer algorithm to generate observations from a homogeneous Poisson processes, or some other renewal process: we simply generate a random sample from the appropriate waiting-time distribution; the i th event time will then be the sum of the first i waiting times. In particular, to generate a homogeneous

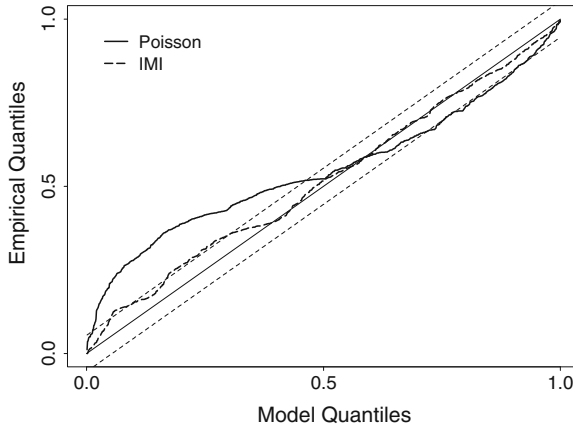


Fig. 19.12 P–P plots of inhomogeneous Poisson and multiplicative IMI models for spike train data from a locust olfactory bulb. For a perfect fit the curve would fall on the diagonal line $y = x$. The data-based (empirical) probabilities deviate substantially from the Poisson model but much less so from the IMI model. When the curve ranges outside the diagonal bands above and below the $y = x$ line, some lack of fit is indicated according to the Kolmogorov-Smirnov test (discussed in Section 10.3.7).

Poisson process with rate λ , we can draw a random sample from an $Exp(\lambda)$ distribution and take the i th event time to be $s_i = \sum_{j=1}^i x_j$.

Generating event times from a general point process is more complicated. One simple approach, based on the Bernoulli approximation, involves partitioning the total time interval into small bins of size Δt : in the k th interval, centered at t_k , we generate an event with probability $p_k = \lambda(t_k|x_{t_k})\Delta t$, where x_{t_k} depends on the history of previously generated events. This works well for small simulation intervals. However, as the total time 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 $\lambda(t|x_t)\Delta t$ is small. In such cases the method becomes very inefficient and thus may take excessive computing time. Alternative approaches generate a relatively small number of i.i.d. observations, and then manipulate them so that the resulting distributions match those of the desired point process.

Thinning To apply this algorithm, the conditional intensity function $\lambda(t|x_t)$ must be bounded by some constant, λ_{\max} . The algorithm follows a two-stage process. In the first stage, a set of candidate event times is generated as a simple Poisson process with a rate λ_{\max} . Because $\lambda_{\max} \geq \lambda(t|x_t)$, these candidate event times occur more frequently than they would for the point process we want to simulate. In the second stage they are “thinned” by removing some of them according to a stochastic scheme. We omit the details. In practice, thinning is typically only used when simulating inhomogeneous Poisson processes with bounded intensity functions.

Time rescaling Another approach to simulating general point processes is based on the time-rescaling theorem. According to the statement of the theorem in Section 19.3.5, the transformed Z_i random variables follow an $Exp(1)$ distribution, with the transformation being based on the integral of the conditional intensity function. This suggests generating a sequence of $Exp(1)$ random variables and then back-transforming to get the desired point process. That idea turns out to work rather well in practice. Here is the algorithm for generating a process on the interval $(0, T]$ with conditional intensity $\lambda(t|x_t)$:

1. Initialize $s_0 = 0$ and $i = 1$.
2. Sample z_i from an $Exp(1)$ distribution.
3. Find s_i as the solution to

$$z_i = \int_{s_{i-1}}^{s_i} \lambda(t|x_t) dt.$$

4. If $s_i > T$ stop.
5. Set $i = i + 1$ and go to 2.

19.3.7 Spectral analysis of point processes requires care.

Because point processes may be considered, approximately, to be binary time series (see Section 19.1.2) it is tempting to treat them as a time series and use spectral methods to find frequency-based components, as in Section 18.3. This is possible, but requires attention to the nature of point processes.

In the first place, spectral analysis applies to stationary time series. To define stationarity (see on p. 515) for a point process we require that the counts $\Delta N_{(t_1, t_2]}$, $\Delta N_{(t_2, t_3]}$, \dots , $\Delta N_{(t_{k-1}, t_k]}$ have the same joint distribution as $\Delta N_{(t_1+h, t_2+h]}$, $\Delta N_{(t_2+h, t_3+h]}$, \dots , $\Delta N_{(t_{k-1}+h, t_k+h]}$ for all h and all $t_1 < t_2 < \dots < t_k$. However, we previously defined point processes only on the positive real line $(0, \infty)$ and for stationarity to make sense the process must be defined on the whole real line $(-\infty, \infty)$. One way to extend a point process to the negative half of the real line is to define the counts to be negative when $t < 0$. For example, suppose we have a homogeneous Poisson process on $(0, \infty)$ with rate λ . Let its counting process representation be $M_1(t)$. Now take another homogeneous Poisson process with rate λ and counting process $M_2(t)$ and define $N(t) = M_1(t)$ for $t > 0$ and $N(t) = -M_2(-t)$ for $t < 0$, and set $N(0) = 0$. Then $N(t)$ becomes the counting process representation of a stationary Poisson process with rate λ .

We now assume that we have counts $\Delta N_{(t_1, t_2]}$ defined for all t and that the resulting point process is stationary. In Section 18.3 the spectral density was defined as the Fourier transform of the autocovariance function. The expectation of a count was given in terms of the marginal intensity in (19.24). In the stationary case the marginal intensity must be time-invariant and therefore equal to a constant λ . We may define a covariance intensity function analogously as

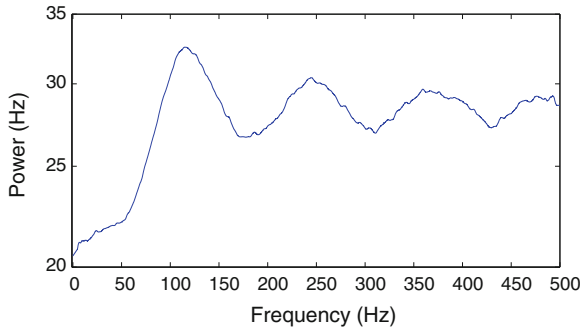


Fig. 19.13 Estimated spectral density from a simulated spike train. The simulated spike train had an average firing rate of roughly 28 Hz, a 5 ms refractory period, and an increased probability of spiking after a previous spike roughly 8 ms in the past. The estimated spectral density does not appear to reflect these properties and is easily misinterpreted.

$$\begin{aligned} \kappa(s, t) &= \lim_{\Delta t \rightarrow 0} \frac{E(\Delta N_{s,s+\Delta t} \Delta N_{t,t+\Delta t}) - E(\Delta N_{s,s+\Delta t})E(\Delta N_{t,t+\Delta t})}{(\Delta t)^2} \\ &= \lim_{\Delta t \rightarrow 0} \frac{E(\Delta N_{s,s+\Delta t} \Delta N_{t,t+\Delta t})}{(\Delta t)^2} - \lambda^2. \end{aligned} \tag{19.38}$$

This holds for $s \neq t$. In the stationary case $\kappa(s, t)$ is a function only of the difference $h = t - s$ so we write $\kappa(h)$ and use (19.38) for $h \neq 0$. For $s = t$ we have, for small Δt (because $\Delta N_{t,t+\Delta t}$ is binary),

$$E(\Delta N_{t,t+\Delta t} \Delta N_{t,t+\Delta t}) = E(\Delta N_{t,t+\Delta t})$$

which implies that the limit in (19.38) vanishes. Instead, we define

$$\kappa(0) = \lim_{\Delta t \rightarrow 0} \frac{V(\Delta N_{t,t+\Delta t})}{\Delta t} = \lambda. \tag{19.39}$$

We therefore must analyze separately¹² the cases $\kappa(0)$ and $\kappa(h)$ when $h \neq 0$. Keeping this in mind, we may now state that the point process spectrum is the Fourier transform of the covariance function. We omit details (see Brillinger 1972).

These technicalities are an indication that point process spectra are likely to behave somewhat differently than continuous spectra. It is possible to apply the discrete Fourier transform to spike train data and then try to interpret the result. Figure 19.13 displays an example of the estimated spectrum of a simulated spike train. Visual inspection of the estimated spectrum shows a dip at low frequencies, a large peak around 120 Hz, and maintained power out to 500 Hz. A naïve interpretation from

¹² These may be combined by writing the covariance function, often called the *complete covariance function* as $\kappa(0)\delta(h) + \kappa(h)$ where $\delta(h)$ is the *Dirac delta function*, which is infinite at 0 and 0 for all other values of h .

this spectrum might presume that this spiking process has no very low frequency firing, tends to fire around 120Hz, but also has considerable high frequency activity, suggesting no refractoriness. However, this interpretation is incorrect. The point process generating this spike train actually has an average firing rate around 28Hz and reflects realistic spiking features including a 5 ms refractory period and an increased probability of firing 8 ms after a previous spike. The error here does not come from the computation of the estimated spectrum, but rather from the naïve interpretation.

We do not pursue further the estimation of point process spectra. Our discussion of Fig. 19.13 is intended to show that point process spectra must be interpreted carefully.

19.4 Additional Derivations

Derivation of Equation (19.9) We start with a lemma.

Lemma The pdf of the i th waiting-time distribution is

$$f_{S_i}(s_i | S_{i-1} = s_{i-1}) = \lambda(s_i) \exp \left\{ - \int_{s_{i-1}}^{s_i} \lambda(t) dt \right\}. \quad (19.40)$$

Proof of the lemma: Note that $\{S_i > s_i | S_{i-1} = s_{i-1}\}$, is equivalent to there being no events in the interval $(s_{i-1}, s_i]$. Therefore, from the definition of a Poisson process on p. 574 together with the Poisson pdf in Eq. (5.3), we have $P(S_i > s_i | S_{i-1} = s_{i-1}) = P(\Delta N_{(s_{i-1}, s_i]} = 0) = \exp \left\{ - \int_{s_{i-1}}^{s_i} \lambda(t) dt \right\}$, and the i th waiting time CDF is therefore $P(S_i \leq s_i | S_{i-1} = s_{i-1}) = 1 - \exp \left\{ - \int_{s_{i-1}}^{s_i} \lambda(t) dt \right\}$. The derivative of the CDF

$$f_{S_i}(s_i | S_{i-1} = s_{i-1}) = \frac{d}{ds_i} \left(1 - \exp \left\{ - \int_{s_{i-1}}^{s_i} \lambda(t) dt \right\} \right)$$

gives the desired pdf. □

Proof of the theorem: We have

$$\begin{aligned} & f_{S_1, \dots, S_{N(T)}}(s_1, \dots, s_n) \\ &= f_{S_1}(s_1) f_{S_2}(s_2 | S_1 = s_1) \cdots f_{S_{N(T)}}(s_n | S_{n-1} = s_{n-1}) \cdot P(\Delta N_{(s_n, T]} = 0). \end{aligned}$$

The factors involving waiting-time densities are given by the lemma. The last factor is

$$P(\Delta N_{(s_n, T]} = 0) = \exp \left(- \int_{s_n}^T \lambda(t) dt \right).$$

Combining these gives the result. □

Derivation of Equation (19.20) We need a lemma, which is analogous to the lemma used in deriving (19.9).

Lemma For an orderly point process with conditional intensity $\lambda(t|H_t)$ on $[0, T]$, the pdf of the i th waiting-time distribution, conditionally on $S_1 = s_1, \dots, S_{i-1} = s_{i-1}$, for $t \in (s_{i-1}, T]$ is

$$f_{S_i|S_1, \dots, S_{i-1}}(s_i|S_1 = s_1, \dots, S_{i-1} = s_{i-1}) = \lambda(s_i|H_t) \exp \left\{ - \int_{s_{i-1}}^{s_i} \lambda(t|H_t) dt \right\}. \tag{19.41}$$

Proof of the lemma: Let X_i be the waiting time for the i th event, conditionally on $S_1 = s_1, \dots, S_{i-1} = s_{i-1}$. For $t > s_{i-1}$ we have $X_i \in (t, t + \Delta t)$ if and only if $\Delta N_{(t, t+\Delta t)} > 0$. Furthermore, if the i th event has not yet occurred at time t we have $H_t = (s_1, \dots, s_{i-1})$. We then have

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \frac{P(X_i \in (t, t + \Delta t)|X_i > t, S_1 = s_1, \dots, S_{i-1} = s_{i-1})}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(\Delta N_{(t, t+\Delta t)} > 0|H_t)}{\Delta t} \end{aligned}$$

and, because the point process is regular, the right-hand side is $\lambda(t|H_t)$. Just as we argued in the case of hazard functions, in Section 3.2.4, the numerator of the left-hand side may be written

$$P(X_i \in (t, t + \Delta t)|X_i > t, H_t) = \frac{F(t + \Delta t|H_t) - F(t|H_t)}{1 - F(t|H_t)}$$

where F is the CDF of the waiting time distribution, conditionally on H_t . Passing to the limit again gives

$$\lim_{\Delta t \rightarrow 0} \frac{P(X_i \in (t, t + \Delta t)|X_i > t, H_t)}{\Delta t} = \frac{f(t|H_t)}{1 - F(t|H_t)}.$$

In other words, just as in the case of a hazard function, the conditional intensity function satisfies

$$\lambda(t|H_t) = \frac{f(t|H_t)}{1 - F(t|H_t)}.$$

Proceeding as in the case of the hazard function we then get the conditional pdf

$$f(t|H_t) = \lambda(t|H_t) e^{-\int_{s_{i-1}}^t \lambda(u|H_u) du}$$

as required. □

Proof of the theorem: The argument follows from the lemma by the same steps as the theorem for inhomogeneous Poisson processes. □

Proof of the time rescaling theorem Note that the transformed waiting times are

$$Z_j = \int_{s_{j-1}}^{s_j} \lambda(u|x_u) du$$

where $s_0 = 0$. Applying the theorem on producing exponential random variables from the probability integral transform (p. 122) to $X_1 = S_1$ with $Z_1 = G(X_1)$ and $G(t) = G_1(t)$ where

$$G_1(t) = \int_0^t \lambda(u|x_u) du,$$

we get $Z_1 \sim \text{Exp}(1)$. Continuing to the next event time and defining $X_2 = S_2 - S_1$ with $Z_2 = G(X_2)$ and $G(t) = G_2(t)$ where

$$G_2(t) = \int_{s_1}^t \lambda(u|x_u) du,$$

we get $Z_2 \sim \text{Exp}(1)$ and, furthermore, this same distribution results regardless of the value of $Z_1 = z_1$. Thus, the conditional density $f_{Z_2|Z_1}(z_2|Z_1 = z_1)$ does not depend on z_1 ; therefore Z_2 is independent of Z_1 . Continuing on, we get $Z_j \sim \text{Exp}(1)$ independently of all Z_i for $i < j$, for all $j = 1, \dots, n$ and for all possible values $n = N(T)$ of the random variable $N(T)$. \square