Jianfeng Feng

Computational Neuroscience: A Comprehensive Approach

CRC PRESS
Boca Raton Ann Arbor London Tokyo

Contents

Chapter 9

Likelihood Methods for Neural Spike Train Data Analysis

Emery N. Brown, Riccardo Barbieri, Uri T. Eden, and Loren M. Frank

Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, U.S., Division of Health Sciences and Technology, Harvard Medical School, Massachusetts Institute of Technology, U.S.

CONTENTS

9.1 Introduction

Computational neuroscience uses mathematical models to study how neural systems represent and transmit information. Although modeling in computational neuroscience spans a range of mathematical approaches, the discipline may be divided approximately into two schools. The first school uses detailed biophysical (Hodgkin and Huxley and their variants) models of individual neurons, networks of neurons or artificial neural network models to study emergent behaviors of neural systems.

1-58488-362-6/04/$0.00+$1.50

The second school, and the one we discuss here, develops signal processing algorithms and statistical methods to analyze the ever-growing volumes of data collected in neuroscience experiments. The growing complexity of neuroscience experiments makes use of appropriate data analysis methods crucial for establishing how reliably specific system properties can be identified from experimental measurements. In particular, careful data analysis is an essential complement to neural network modeling; it allows validation of neural network model predictions in addition to feeding back biologically relevant constraints and parameter values for further analytic and simulation studies. Neuroscience experiments and neural spike train data have special features that present new, exciting challenges for statistical research.

Neuroscience data analyses as well as research on new data analysis methods should exploit established statistical paradigms wherever possible. Several standard statistical procedures, widely used in other fields of science have been slow to find their way into mainstream application in neuroscience data analysis. One such set of procedures are those based on the likelihood principle [9, 31]. The likelihood function is a central tool in statistical theory and modeling, typically based on a parametric model of an experimental data set. The likelihood is formulated by deriving the joint distribution of the data, and then viewing this joint distribution as a function of the model parameters with the data fixed. This function serves as a criterion function for estimating the model parameters, assessing goodness-of-fit, constructing confidence statements, and eventually, for making inferences about the particular problem under study. The several optimality properties of the likelihood approach is one of the main reasons this paradigm is central to statistical theory and data analysis. Neural spike trains are point process measurements. Therefore, to help better acquaint neuroscientists with likelihood-based methods, we review the likelihood paradigm for point process observations.

The remainder of this chapter is organized as follows. In Section 9.2, we show how any point process model may be characterized in terms of its conditional intensity function. The conditional intensity function is a history-dependent generalization of the rate function for a Poisson process. It provides a canonical representation of the stochastic properties of a neural spike train. We use the conditional intensity function to derive the joint probability density of the neural spike train and hence, its likelihood function. We next review briefly the optimality properties of the likelihood approach and we show how the conditional intensity function may be used to derive goodness-of-fit tests based on the time-rescaling theorem. In Section 9.3 we apply our likelihood methods in three actual data analyses. In the first example we compare the fits of exponential, gamma and inverse Gaussian interspike interval distribution models to a spike train time-series from a retinal ganglion neuron. In the second example, we use likelihood and non-likelihood based methods to analyze the spatial receptive fields of a hippocampal neuron recorded while a rat executes a behavioral task on a linear track. In the third example, we show how the likelihood function may be used to construct a criterion function for adaptive estimation that makes it possible to track plasticity in a neural receptive field on a millisecond time-scale. We illustrate the method by performing a dynamic analysis of the spatial receptive field of a hippocampal neuron from the same linear track experiment studied in the second

example. Section 9.4 presents a set of conclusions.

9.2 Theory

9.2.1 The conditional intensity function and interspike interval probability density

As mentioned in Section 9.1, the key to deriving the likelihood function for a parametric model of a neural spike train is defining the joint probability density. The joint probability density of a neural spike train can be characterized in terms of the conditional intensity function. Therefore, we first derive the conditional intensity function for a point process and review some of its properties.

Let $(0,T]$ denote the observation interval and let $0 < u_1 < u_2 < \cdots < u_{J-1} < u_J \leq T$ be a set of J spike time measurements. For $t \in (0,T]$ let $N_{0:t}$ be the sample path of the point process over $(0,t]$. It is defined as the event $N_{0:t} = \{0 < u_1 < u_2 \cdots < u_j \leq t \cap N(t) = j\}$, where $N(t)$ is the number of spikes in $(0,t]$ and $j \leq J$. The sample path is a right continuous function that jumps 1 at the spike times and is constant otherwise [11, 33]. The function $N_{0:t}$ tracks the location and number of spikes in $(0,t]$ and hence, contains all the information in the sequence of spike times (Figure 9.1A).

We define the conditional intensity function for $t \in (0,T]$ as

$$\lambda(t|H_t) = \lim_{\triangle \to 0} \frac{Pr(N(t+\triangle) - N(t) = 1|H_t)}{\triangle} \tag{9.1}$$

where H_t is the history of the sample path and of any covariates up to time t. In general $\lambda(t|H_t)$ depends on the history of the spike train and therefore, it is also termed the stochastic intensity [11]. In survival analysis, the conditional intensity function is called the hazard function [20]. This is because the hazard function can be used to define the probability of an event in the interval $[t, t+\triangle)$ given that there has not been an event up to t. It follows that $\lambda(t|H_t)$ can be defined in terms of the inter-event or spike time probability density at time t, $p(t|H_t)$, as

$$\lambda(t|H_t) = \frac{p(t|H_t)}{1 - \int_0^t p(u|H_u)du} \tag{9.2}$$

We gain insight into the definition of the conditional intensity function in Equstion (9.1) by considering the following heuristic derivation of Equation (9.2) based on the definition of the hazard function. We compute explicitly the probability of the event,

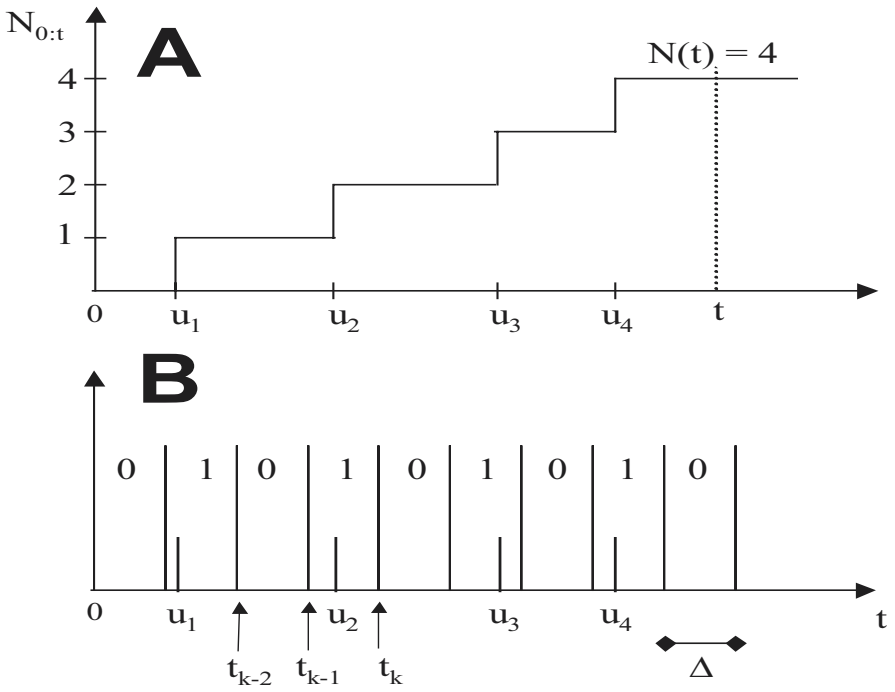

Figure 9.1

A. The construction of the sample path $N_{0:t}$ from the spike times $u_1, \cdots, u_4$. At time t, $N_{0:t} = \{u_1, u_2, u_3, u_4 \cap N(t) = 4\}$ B. The discretization of the time axis to evaluate to evaluate the probability of each spike occurrence or non-occurrence as a local Bernoulli process. By Equation (9.10) the probability of the event u_2, i.e., a 1 between t_{k-1} and t_k, is $\lambda(t_k|H_k)\triangle$ whereas the probability of the event immediately prior to u_2, i.e., a 0 between t_{k-2} and t_{k-1}, is $1-\lambda(t_{k-1}|H_{k-1})\triangle$. In this plot we have taken $\triangle_k = \triangle$ for all $k = 1, \cdots, K$.

a spike in $[t,t+\triangle)$ given H_t and that there has been no spike in $(0,t)$. That is,

$$\begin{aligned} Pr(u \in [t,t+\triangle)|u>t,H_t) &= \frac{Pr(u \in [t,t+\triangle) \cap u>t|H_t)}{Pr(u>t|H_t)} \\ &= \frac{Pr(u \in [t,t+\triangle)|H_t)}{Pr(u>t|H_t)} \\ &= \frac{\int_t^{t+\triangle} p(u|H_u)du}{1-\int_0^t p(u|H_u)du} \\ &= \frac{p(t|H_t)\triangle}{1-\int_0^t p(u|H_u)du} + o(\triangle) \\ &= \lambda(t|H_t)\triangle + o(\triangle) \end{aligned} \tag{9.3}$$

where $o(\triangle)$ refers to all events of order smaller than $\triangle$, such as two or more events (spikes) occurring in an arbitrarily small interval. This establishes Equation (9.2). The power of the conditional intensity function is that if it can be defined as Equation (9.3) suggests then, it completely characterizes the stochastic structure of the spike train. In any time interval $[t,t+\triangle), \lambda(t|H_t)\triangle$ defines the probability of a spike given the history up to time t. If the spike train is an inhomogeneous Poisson process then, $\lambda(t|H_t) = \lambda(t)$ becomes the Poisson rate function. Thus, the conditional intensity function (Equation (9.1)) is a history-dependent rate function that generalizes the definition of the Poisson rate. Similarly, Equation (9.1) is also a generalization of the hazard function for renewal processes [15, 20].

We can write

$$\lambda(t|H_t) = -\frac{d\left[\log[1-\int_0^t p(u|H_u)du]\right]}{dt} \tag{9.4}$$

or on integrating we have

$$-\int_0^t \lambda(u|H_u)du = \log\left[1-\int_0^t p(u|H_u)du\right]. \tag{9.5}$$

Finally, exponentiating yields

$$\exp\left\{-\int_0^t \lambda(u|H_u)du\right\} = 1-\int_0^t p(u|H_u)du. \tag{9.6}$$

Therefore, by Equations. (9.2) and (9.6) we have

$$p(t|H_t) = \lambda(t|H_t)\exp\left\{-\int_0^t \lambda(u|H_u)du\right\}. \tag{9.7}$$

Together Equations (9.2) and (9.7) show that given the conditional intensity function the interspike interval probability density is specified and vice versa. Hence,

defining one completely defines the other. This relation between the conditional intensity or hazard function and the inter-event time probability density is well known in survival analysis and renewal theory [15, 20]. Equations (9.2) and (9.7) show that it holds for a general point process model. This relation is exploited in the data analysis examples we discuss.

9.2.2 The likelihood function of a point process model

The likelihood of a neural spike train, like that of any statistical model, is defined by finding the joint probability density of the data. We show in the next proposition that the joint probability of any point process is easy to derive from the conditional intensity function.

Proposition 1. Given $0 < u_1 < u_2 < \cdots < u_J < T$, a set of neural spike train measurements, the sample path probability density of this neural spike train, i.e. the probability density of exactly these J events in $(0,T]$, is

$$\begin{aligned} p(N_{0:T}) &= \prod_{j=1}^{J} \lambda(u_j|H_{u_j}) \exp\left\{-\int_0^T \lambda(u|H_u)du\right\} \\ &= \exp\left\{\int_0^T \log\lambda(u|H_u)dN(u) - \int_0^T \lambda(u|H_u)du\right\}. \end{aligned} \tag{9.8}$$

Proof. Let $\{t_k\}_{k=1}^K$ be a partition of the observation interval $(0,T]$. Take $\triangle_k = t_k - t_{k-1}$, where $t_0 = 0$. Assume that the partition is sufficiently fine so that there is at most one spike in any $(t_{k-1},t_k]$. For a neural spike train choosing $\triangle_k \leq 1$ msec would suffice. We define $dN(k) = 1$ if there is a spike in $(t_{k-1},t_k]$ and 0 otherwise, and the events

$$\begin{aligned} A_k &= \{\text{spike in } (t_{k-1},t_k]|H_k\} \\ E_k &= \{A_k\}^{dN(k)}\{A_k^c\}^{1-dN(k)} \\ H_k &= \left\{\cap_{j=1}^{k-1} E_j\right\} \end{aligned} \tag{9.9}$$

for $k = 1,\cdots,K$. In any interval $(t_{k-1},t_k]$ we have (Figure 9.1B)

$$\begin{aligned} Pr(E_k) &= \lambda(t_k|H_k)\triangle_k + o(\triangle_k) \\ Pr(E_k^c) &= 1 - \lambda(t_k|H_k)\triangle_k + o(\triangle_k). \end{aligned} \tag{9.10}$$

By construction of the partition we must have $u_j \in (t_{k_j-1},t_{k_j}], j = 1,\cdots,J$ for a subset of the intervals satisfying $k_1 < k_2 \cdots < k_J$. The remaining $K - J$ intervals have no spikes. The spike events form a sequence of correlated Bernoulli trials. It follows from Equation (9.10) and the Lemma in the Appendix, that given the partition, the

probability of exactly J events in $(0,T]$ may be computed as

$$\begin{aligned}
p(N_{0:T})\prod_{j=1}^{J}\triangle_{k_j} &= p(u_j \in (t_{k_j-1},t_{k_j}], j=1,\cdots,J \cap N(T)=J)\prod_{j=1}^{J}\triangle_{k_j} \\
&= Pr(\cap_{k=1}^{K}E_k) \\
&= \prod_{k=2}^{K} Pr(E_k|\cap_{j=1}^{k-1}E_j)Pr(E_1) \\
&= \prod_{k=1}^{K}[\lambda(t_k|H_k)\triangle_k]^{dN(t_k)}[1-\lambda(t_k|H_k)\triangle_k]^{1-dN(t_k)}+o(\triangle_*) \\
&= \prod_{j=1}^{J}[\lambda(t_{k_j}|H_{k_j})\triangle_{k_j}]^{dN(t_{k_j})}\prod_{l\neq k_j}[1-\lambda(t_l|H_l)\triangle_l]^{1-dN(t_l)}+o(\triangle_*) \\
&= \prod_{j=1}^{J}[\lambda(t_{k_j}|H_{k_j})\triangle_{k_j}]^{dN(t_{k_j})}\prod_{l\neq k_j}\exp\{-\lambda(t_l|H_l)\triangle_l\}+o(\triangle_*) \\
&= \exp\left\{\sum_{j=1}^{J}\log\lambda(t_{k_j}|H_{k_j})dN(t_{k_j})-\sum_{l\neq k_j}\lambda(t_l|H_l)\triangle_l\right\} \\
&\quad\cdot\exp\left\{\sum_{j=1}^{J}\log\triangle_{k_j}\right\}+o(\triangle_*)
\end{aligned} \tag{9.11}$$

where, because the $\triangle_k$ are small, we have used the approximation $[1-\lambda(k)\triangle_k]\approx \exp\{-\lambda(k)\triangle_k\}$ and $\triangle_*=\max_k\triangle_k$. It follows that the probability density of exactly these J spikes in $(0,T]$ is

$$\begin{aligned}
p(N_{0:T}) &= \lim_{\triangle_*\to 0}\left[\frac{\exp\left\{\sum_{j=1}^{J}\log\lambda(t_{k_j}|H_{k_j})dN(t_{k_j})-\sum_{l\neq k_j}\lambda(t_l|H_l)\triangle_l\right\}}{\prod_{j=1}^{J}\triangle_j}\right. \\
&\quad\left.\cdot\exp\left\{\sum_{j=1}^{J}\log\triangle_{k_j}\right\}+\frac{o(\triangle_*)}{\prod_{j=1}^{J}\triangle_j}\right] \\
&= \exp\left\{\int_0^T\log\lambda(u|H_u)dN(u)-\int_0^T\lambda(u|H_u)du\right\}. \qquad \textbf{Q.E.D.}
\end{aligned} \tag{9.12}$$

Proposition 1 shows that the joint probability density of a spike train process can be written in a canonical form in terms of the conditional intensity function [3, 8, 11]. That is, when formulated in terms of the conditional intensity function, all point process likelihoods have the form given in Equation (9.8). The approximate probability density expressed in terms of the conditional intensity function (Equation (9.11d)) was given in [5]. The proof of Proposition 1 follows the derivation in [1]. The insight provided by this proof is that correct discretization for computing the

local probability of a spike event is given by the conditional intensity function. An alternative derivation of Equation (9.8) can be obtained directly using Equation (9.7) [3].

If the probability density in Equation (9.8) depends on an unknown q-dimensional parameter θ to be estimated then, Equation (9.8) viewed as a function of θ given $N_{0:T}$ is the likelihood function defined as

$$\begin{aligned} L(\theta|N_{0:T}) &= p(N_{0:T}|\theta) \\ &= \exp\left\{\int_0^T \log\lambda(u)|H_u,\theta)dN(u) - \int_0^T \lambda(u|H_u)du\right\}. \end{aligned} \tag{9.13}$$

The logarithm of Equation 9.13 is the log likelihood function defined as

$$\log L(\theta|N_{0:T}) = \int_0^T l_u(\theta)du \tag{9.14}$$

where $l_t(\theta)$ is the integrand in Equation (9.14) or the *instantaneous* log likelihood defined as

$$l_t(\theta) = \log[\lambda(t|H_t,\theta)]\frac{dN(t)}{dt} - \lambda(t|H_t,\theta). \tag{9.15}$$

Given a model for the spike train, defined either in terms of the conditional intensity function or the interspike interval probability density, the likelihood is an objective quantity that offers a measure of *rational belief* [9, 31]. Specifically, the likelihood function measures the relative preference for the values of the parameter given the observed data $N_{0:T}$. Similarly, the instantaneous log likelihood in Equation (9.15) may be viewed as measuring the instantaneous accrual of *information* from the spike train about the parameter θ. We will illustrate in the applications in Section 9.3 how methods to analyze neural spike train data may be developed using the likelihood function. In particular, we will use the instantaneous log likelihood as the criterion function in the point process adaptive filter algorithm presented in Section 9.3.3.

9.2.3 Summarizing the likelihood function: maximum likelihood estimation and Fisher information

If the likelihood is a one or two-dimensional function it can be plotted and completely analyzed for its information content about the model parameter θ. When the dimension of θ is greater than 2, a complete analysis of the likelihood function by graphical methods is not possible. Therefore, it is necessary to summarize this function. The most common way to summarize the likelihood is to compute the maximum likelihood estimate of the parameter θ. That is, we find the value of this parameter that is most likely given the data. This corresponds to the value of θ that makes Equation (9.13) or equivalently, Equation (9.14) as large as possible. We define the maximum likelihood estimate $\hat{\theta}$ as

$$\hat{\theta} = \arg\max_\theta L(\theta|N_{0:T}) = \arg\max_\theta \log L(\theta|N_{0:T}). \tag{9.16}$$

With the exception of certain elementary models the value of θ that maximizes Equation (9.16) has to be computed numerically. In most multidimensional problems it is

difficult to insure that the numerical analysis will yield a global maximum. Most often a local rather than a global estimate is obtained. Several different starting values of parameters should be used in the numerical optimization procedure to increase the probability of obtaining a global maximum of the likelihood.

A second standard statistic computed to summarize the likelihood is the Fisher Information. The Fisher Information is defined as

$$I(\theta) = -E[\nabla^2 \log L(\theta|N_{0:T})], \tag{9.17}$$

where ∇^2 is the Hessian of the log likelihood with respect to θ and E denotes the expectation taken with respect to $p(N_{0:T}|\theta)$. The Fisher Information matrix can be used to measure the uncertainty in the maximum likelihood estimate. This is because under not too stringent regularity conditions, the asymptotic (large sample) distribution of the maximum likelihood estimate $\hat{\theta}$ is the Gaussian distribution whose mean is the true value of the parameter θ, and whose covariance matrix is $I(\theta)^{-1}$ [9, 31]. Because θ is unknown we evaluate $I(\theta)$ as $I(\hat{\theta})$ or $I(\hat{\theta})_{N_{0:T}} = -\nabla^2 \log L(\hat{\theta}|N_{0:T})$ where the latter is termed the observed Fisher information. Under the Gaussian approximation to the asymptotic distribution of the maximum likelihood estimate, the Fisher information may be used to compute confidence intervals for the components of the true parameter vector given by

$$\hat{\theta}_i \pm z_{1-\alpha/2}[I(\hat{\theta}_i)]^{-1/2}, \tag{9.18}$$

where $\hat{\theta}_i$ is the i^{th} component of $\hat{\theta}$ for $i = 1, \cdots, q$ and $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard Gaussian distribution.

Another way of viewing the maximum likelihood estimate along with the Fisher information is as a means of constructing a Gaussian approximation to the likelihood function. By expanding $\log L(\theta|N_{0:T})$ in a Taylor series about $\hat{\theta}$ we obtain the following Gaussian approximation to $L(\theta|N_{0:T})$ namely,

$$L(\theta|N_{0:T}) \sim [(2\pi)^q |I(\hat{\theta})|]^{-1/2} \exp\left\{ -\frac{(\theta-\hat{\theta})^T I(\hat{\theta})^{-1}(\theta-\hat{\theta})}{2} \right\}. \tag{9.19}$$

While Equation (9.19) is functionally equivalent to the statement that the maximum likelihood estimate has an approximate Gaussian probability density, this equation has a Bayesian interpretation. This is because in the classical or frequentist statement that the asymptotic distribution of the maximum likelihood estimate is Gaussian, $\hat{\theta}$ is a random variable and θ, the true parameter value is a fixed quantity. In Equation (9.19) $\hat{\theta}$ and $I(\hat{\theta})$ are fixed quantities and θ is a random variable [31, 35].

9.2.4 Properties of maximum likelihood estimates

One of the most compelling reasons to use maximum likelihood estimation in neural spike train data analyses is that for a broad range of models, these estimates have other important optimality properties in addition to being asymptotically Gaussian.

First, there is consistency which states that the sequence of maximum likelihood estimates converges in probability (or more strongly almost surely) to the true value as the sample size increases. Second, the convergence in probability of the estimates means that they are asymptotically unbiased. That is, the expected value of the estimate $\hat{\theta}$ is θ as the sample size increases. For some models and some parameters, unbiasedness is a finite sample property. The third property is invariance. That is, if $\hat{\theta}$ is the maximum likelihood estimate of θ, then $\tau(\hat{\theta})$ is the maximum likelihood estimate of $\tau(\theta)$. Finally, the maximum likelihood estimates are asymptotically efficient in that as the sample size increases, the variance of the maximum likelihood estimate achieves the Cramer-Rao lower bound. This lower bound defines the smallest variance that an unbiased or asymptotically unbiased estimate can achieve. Like unbiasedness, efficiency for some models and some parameters is achieved in a finite sample. Detailed discussions of these properties are given in [9, 31].

9.2.5 Model selection and model goodness-of-fit

In many data analyses it is necessary to compare a set of models for a given neural spike train. For models fit by maximum likelihood, a well-known approach to model selection is the Akaike's Information Criterion (AIC) [31]. The criterion is defined as

$$AIC = -2\log L(\hat{\theta}|N_{0:T}) + 2q, \tag{9.20}$$

where q is the dimension of the parameter vector θ. The AIC measures the trade-off between how well a given model fits the data and the number of model parameters needed to achieve this fit. The fit of the model is measured by the value of $-2x$ the maximized likelihood and the cost of the number of fitted parameters is measured by $2q$. Under this formulation, i.e. considering the negative of the maximized likelihood, the model that describes the data best in terms of this trade-off will have the smallest AIC.

Evaluating model goodness-of-fit, i.e., measuring quantitatively the agreement between a proposed model and a spike train data series, is a more challenging problem than for models of continuous-valued processes. Standard distance discrepancy measures applied in continuous data analyses, such as the average sum of squared deviations between recorded data values and estimated values from the model, cannot be directly computed for point process data. Berman [4] and Ogata [28] developed transformations that, under a given model, convert point processes like spike trains into continuous measures in order to assess model goodness-of-fit. One of the transformations is based on the time-rescaling theorem.

A form of the time-rescaling theorem is well known in elementary probability theory. It states that any inhomogeneous Poisson process may be rescaled or transformed into a homogeneous Poisson process with a unit rate [36]. The inverse transformation is a standard method for simulating an inhomogeneous Poisson process from a constant rate (homogeneous) Poisson process. Meyer [26] and Papangelou [30] established the general time-rescaling theorem, which states that any point process with an integrable rate function may be rescaled into a Poisson process with a

unit rate. Berman and Ogata derived their transformations by applying the general form of the theorem. An elementary proof of the time-rescaling theorem is given in [8].

We use the time-rescaling theorem to construct goodness-of-fit tests for a neural spike data model. Once a model has been fit to a spike train data series we can compute from its estimated conditional intensity the rescaled times

$$\tau_j = \int_{u_{j-1}}^{u_j} \lambda(u|H_u)du, \tag{9.21}$$

for $j = 1,\cdots,J$. If the model is correct then, according to the theorem, the τ_js are independent exponential random variables with mean 1. If we make the further transformation

$$z_j = 1 - \exp(-\tau_j), \tag{9.22}$$

then z_js are independent uniform random variables on the interval [0,1). Because the transformations in Eqs. (9.21) and (9.22) are both one-to-one, any statistical assessment that measures agreement between the z_js and a uniform distribution directly evaluates how well the original model agrees with the spike train data. We use Kolmogorov-Smirnov tests to make this evaluation [8].

To construct the Kolmogorov-Smirnov test we order the z_js from smallest to largest, denoting the ordered values as $z_j s$ and then plot the values of the cumulative distribution function of the uniform density defined as $b_j = (j-1/2)/J$ for $j = 1,\cdots,J$ against the z_js. If the model is correct, then the points should lie on a 45^o line. Confidence bounds for the degree of agreement between the models and the data may be constructed using the distribution of the Kolmogorov-Smirnov statistic [19]. For moderate to large sample sizes the 95% confidence bounds are well approximated as $b_j \pm 1.36/J^{1/2}$ [19]. We term these plots Kolmogorov-Smirnov (KS) plots.

9.3 Applications

9.3.1 An analysis of the spiking activity of a retinal neuron

In this first example we study a spike train data series from a goldfish retinal ganglion cell neuron recorded in vitro (Figure 9.2). The data are 975 spikes recorded over 30 seconds from neuron 78 in [18]. They were provided by Dr. Satish Iyengar from experiments originally conducted by Dr. Michael Levine at the University of Illinois [22, 23]. The retinae were removed from the goldfish and maintained in a flow of moist oxygen. Recordings of retina ganglion cells were made with an extracellular microelectrode under constant illumination.

The plot of the spikes from this neuron (Figure 9.2) reveals a collection of short and long interspike intervals (ISI). To analyze these data we consider three ISI probability models: the gamma, exponential and inverse Gaussian probability densities.

The gamma probability density is a probability model frequently used to describe renewal processes. It is the ISI probability model obtained from a simple stochastic integrate-and-fire model in which the inputs to the neuron are Poisson with a constant rate [37]. It has the exponential probability density, the interspike interval model associated with a simple Poisson process, as a special case. The inverse Gaussian probability density is another renewal process model that can be derived from a stochastic integrate-and-fire model in which the membrane voltage is represented as a random walk with drift [37]. This model was first suggested by Schroedinger [32] and was first applied in spike train data analyses by Gerstein and Mandelbrot [14]. Because the gamma and inverse Gaussian ISI probability densities can be derived from elementary stochastic integrate-and-fire models, these probability densities suggest more plausible points of departure for constructing statistical models of neural spike trains than the Poisson process.

The gamma and inverse Gaussian probability densities are, respectively,

$$p_1(w_j|\theta) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} w_j^{\alpha-1} \exp\left\{-\lambda w_j\right\}, \tag{9.23}$$

where $\theta = (\alpha, \lambda), \alpha > 0, \lambda > 0$,

$$p_2(w_j|\theta) = \left(\frac{\lambda}{2\pi w_j^3}\right)^{1/2} \exp\left\{-\frac{1}{2}\frac{\lambda(w_j-\mu)^2}{\mu^2 w_j}\right\}, \tag{9.24}$$

where $\theta = (\mu, \lambda), \mu > 0, \lambda > 0$ and $w_j = u_j - u_{j-1}$ for $j = 1, \cdots, J$. For the gamma (inverse Gaussian) model $\alpha(\mu)$ is the location parameter and $\lambda(\lambda)$ is the scale parameter. If $\alpha = 1$ then the gamma model is the exponential probability density. The mean and variance of the gamma model are respectively $\alpha\lambda^{-1}$ and $\alpha\lambda^{-2}$ whereas the mean and variance of the inverse Gaussian model are respectively μ and $\mu^3\lambda^{-1}$. Fitting these models to the spike train data requires construction of the likelihoods and estimation of θ for the three models. By our results in Section 9.2, the log likelihood can be represented either in terms of the conditional intensity or the ISI probability model. Here we use the latter. Given the set of ISIs, $w = (w_1, \cdots, w_J)$, then, under the assumption that the ISIs are independent, the likelihood functions for the two models are respectively

$$L_1(\theta|w) = \prod_{j=1}^{J} p_1(w|\theta) = \left[\frac{\lambda^{\alpha}}{\Gamma(\alpha)}\right]^J \prod_{j=1}^{J} w_j^{\alpha-1} \exp\left\{-\lambda w_j\right\} \tag{9.25}$$

$$L_2(\theta|w) = \prod_{j=1}^{J} p_2(w|\theta) = \prod_{j=1}^{J} \left(\frac{\lambda}{2\pi w_j^3}\right)^{1/2} \exp\left\{-\frac{1}{2}\frac{\lambda(w_j-\mu)^2}{\mu^2 w_j}\right\}. \tag{9.26}$$

The maximum likelihood estimate of θ for the gamma model cannot be computed in

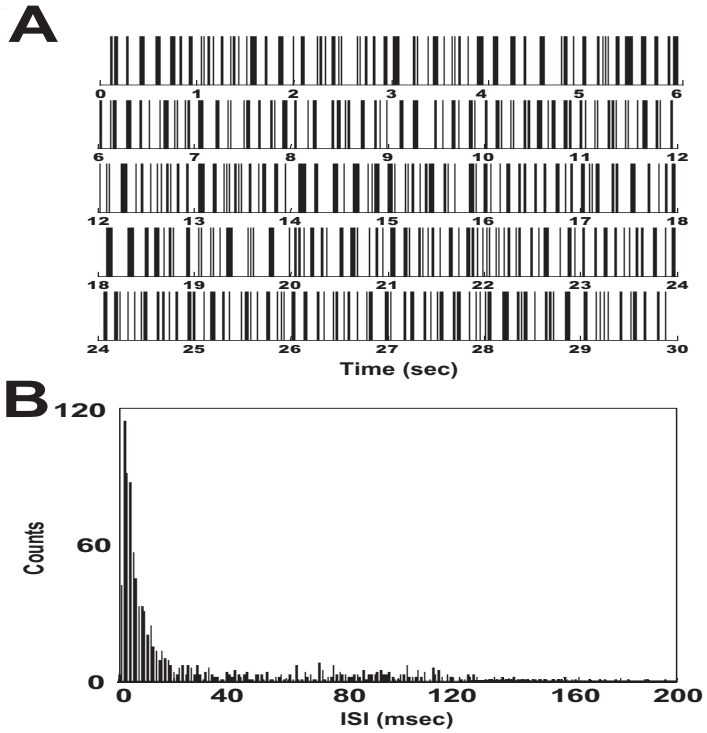

Figure 9.2

A. Thirty seconds of spike times from a retinal ganglion neuron recorded *in vitro* under constant illumination. There is an obvious mixture of short and long interspike intervals. B. Interspike interval histogram for the neural spike train in A. While most of the spikes occur between 3 to 40 msec, there are many intervals longer than 70 msec.

closed form, but rather numerically as the solution to the equations

$$\hat{\lambda} = \frac{\hat{\alpha}}{\bar{w}} \tag{9.27}$$

$$J \log \Gamma(\hat{\alpha}) = J\alpha \log\left(\frac{\hat{\alpha}}{\bar{w}}\right) + (\hat{\alpha} - 1)\sum_{j=1}^{J} \log w_j \tag{9.28}$$

where $\bar{w} = J^{-1}\sum_{j=1}^{J} w_j$ and the $\hat{}$ denotes the estimate. Equations (9.27) and (9.28) are obtained by differentiating Equation (9.25) with respect to α and λ and setting the derivatives equal to zero. It follows from Equation (9.17) that the Fisher Information matrix is

$$I(\theta) = J \begin{bmatrix} \dfrac{\Gamma(\alpha)\Gamma''(\alpha) - \Gamma'(\alpha)\Gamma'(\alpha)}{\Gamma^2(\alpha)} & -\dfrac{1}{\lambda} \\ -\dfrac{1}{\lambda} & \dfrac{\alpha}{\lambda^2} \end{bmatrix}. \tag{9.29}$$

Similarly, differentiating Equation (9.26) with respect to μ and λ and setting the derivatives equal to zero yields as the maximum likelihood estimate of the inverse Gaussian model parameters

$$\hat{\mu} = J^{-1}\sum_{j=1}^{J} w_j \tag{9.30}$$

$$\hat{\lambda}^{-1} = J^{-1}\sum_{j=1}^{J} (w_j^{-1} - \hat{\mu}^{-1}). \tag{9.31}$$

On evaluating Equation (9.17) for this model we find that the Fisher Information matrix is

$$I(\theta) = J \begin{bmatrix} \lambda\mu^{-3} & 0 \\ 0 & (2\lambda^2)^{-1} \end{bmatrix}. \tag{9.32}$$

We compare the fits to the spike train of the exponential, gamma and inverse Gaussian models by comparing the model probability density estimate for each to the normalized histogram of the ISIs (Figure 9.3). The exponential model underpredicts the number of short ISIs ($<$ 10 msec), overpredicts the number of intermediate ISIs (10 to 50 msec) (Figure 9.3B), and underpredicts the number of long ISIs ($>$ 120 msec), (Figure 9.3C). While the gamma model underpredicts the number of short ISIs, ($<$ 10 msec) more than the exponential model, it predicts well the number of intermediate ISIs (10 to 50 msec) (Figure 9.3B), and also underpredicts the number of long ISIs ($>$ 120 msec), (Figure 9.3C). Because the gamma model estimate of α is $\hat{\alpha} = 0.81$ (Table 1), the mode of this probability density lies at zero where the probability density is infinite. Zero lies outside the domain of the probability density as it is defined only for ISIs that are strictly positive. This explains the monotonically decreasing shape of this probability density seen in the plot. Although not completely accurate, the inverse Gaussian model gives the best fit to the short ISIs (Figure 9.3B). The inverse Gaussian model also describes well the intermediate ISIs

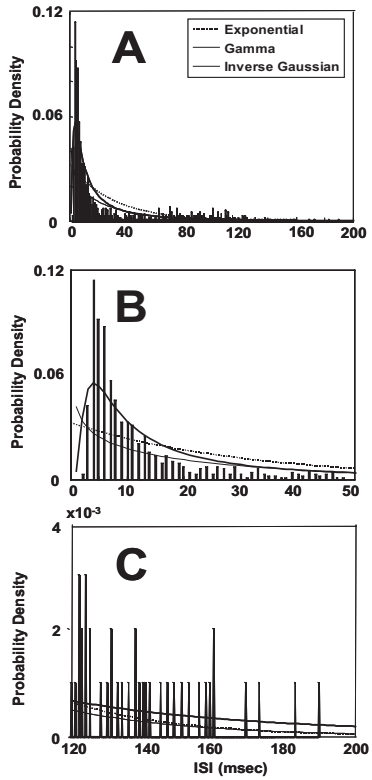

Figure 9.3

A. Maximum likelihood fits of the exponential (dotted line), gamma (solid line), inverse Gaussian (solid bold line) models to the retinal neuron spike trains in Figure 9.2A displayed superimposed on a normalized version of the interspike interval histogram in Figure 9.2B. B. Enlargement from (A) of the interspike interval histogram from 0 to 50 msec to display better the data, and the three model fits over this range. C. Enlargement from (A) of the interspike interval histogram from 120 to 200 msec and the three model fits.

from 25 to 50 msec (Figure 9.3B) and of the three models, underpredicts the long ISIs the least (Figure 9.3C).

Because of Equation (9.2), specifying the spike time probability density is equivalent to specifying the conditional intensity function. From Equation (9.2) and the invariance of the maximum likelihood estimate discussed in Section 9.2.5, it follows that if $\hat{\theta}$ denotes the maximum likelihood estimate of θ then the maximum likelihood estimate of the conditional intensity function for each model can be computed from Equation (9.2) as

$$\lambda(t|H_t,\hat{\theta}) = \frac{p(t|u_{N(t)},\hat{\theta})}{1-\int_{u_{N(t)}}^{t} p(u|u_{N(t)},\hat{\theta})du}, \tag{9.33}$$

for $t > u_{N(t)}$ where $u_{N(t)}$ is the time of the last spike prior to t. The estimated conditional intensity from each model may be used in the time-rescaling theorem to assess model goodness-of-fit as described in Section 9.2.5.

An important advantage of the KS plot is that it allows us to visualize the goodness-of-fit of the three models without having to discretize the data into a histogram (Figure 9.4). While the KS plot for neither of the three models lies entirely within the 95% confidence bounds, the inverse Gaussian model is closer to the confidence bounds over the entire range of the data. These plots also show that the gamma model gives a better fit to the data than the exponential model.

The AIC and KS distance are consistent with the KS plots (Table 1). The inverse Gaussian model has the smallest AIC and KS distance, followed by the gamma and exponential models in that order for both. The approximate 95% confidence interval for each model parameter was computed from maximum likelihood estimates of the parameters (Equations (9.27), (9.28), (9.30), (9.31)) and the estimated Fisher information matrices (Equations (9.29), (9.32)) using Equation (9.18). Because none of the 95% confidence intervals includes zero, all parameter estimates for all three models are significantly different from zero. While all three models estimate the mean ISI as 30.73 msec, the standard deviation estimate from the inverse Gaussian model of 49.0 msec is more realistic given the large number of long ISIs (Figure 9.2B).

In summary, the inverse Gaussian model gives the best overall fit to the retinal ganglion spike train series. This finding is consistent with the results of [18] who showed that the generalized inverse Gaussian model describes the data better than a lognormal model. Our inverse Gaussian model is the special case of their generalized inverse Gaussian model in which the index parameter of the Bessel function in their normalizing constant equals -0.5. In their analyses Iyengar and Liao estimated the index parameter for this neuron to be -0.76. The KS plots suggests that the model fits can be further improved. The plot of the spike train data series (Figure 9.2) suggests that the fit may be improved by considering an ISI model that would specifically represent the obvious propensity of this neuron to burst as well as produce long ISIs in a history-dependent manner. Such a model could be derived as the mixture model that Iyengar and Liao [18] suggest as a way of improve their generalized inverse Gaussian fits.

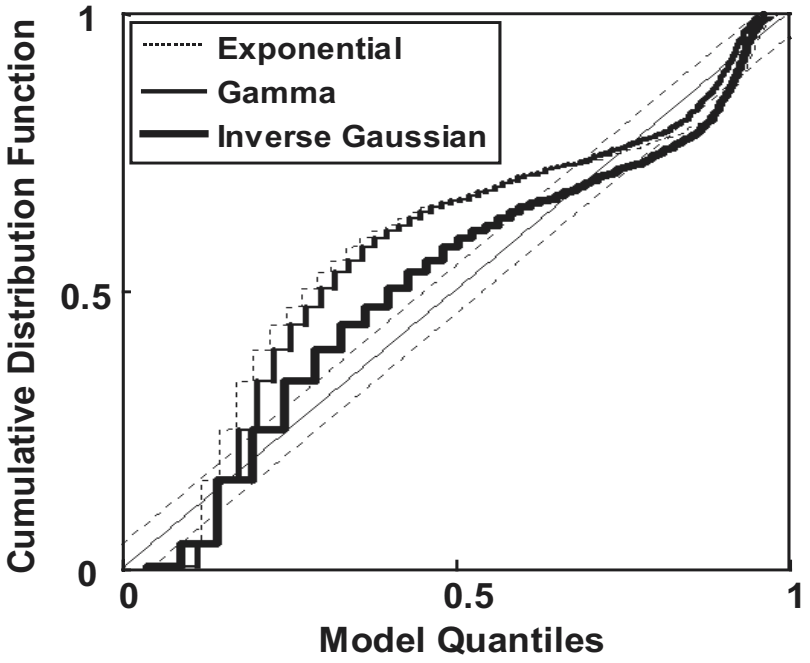

Figure 9.4

Kolmogorov-Smirnov plots for the fits of the exponential (dotted line), gamma (solid line), and inverse Gaussian (solid bold line) models to the neural spike train in Figure 9.2. The parallel diagonal lines are the 95% confidence bounds for the degree of agreement between the models and the spike train data. By this criterion, statistically acceptable agreement between a model and the data would be seen if the KS plot for that model fell entirely within the confidence bounds.

	Exponential	Gamma		Inverse Gaussian	
	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\lambda}$
$\hat{\theta}$	0.0325	0.805	0.262	30.76	12.1
CI	[0.0283 0.0367]	[0.678 0.931]	[0.208 0.316]	[24.46 37.06]	[9.9 14.3]
ISI	30.77±30.77	30.73± 34.74		30.76± 49.0	
AIC	8598	8567		8174	
KS	0.233	0.2171		0.1063	

Table 1: The row is the maximum likelihood estimate $\hat{\theta}$. CI (95% confidence interval for the parameter); ISI (interspike interval mean and SD); AIC (Akaike's Information Criterion); KS (Kolmogorov-Smirnov statistic).

9.3.2 An analysis of hippocampal place-specific firing activity

As a second example of applying likelihood methods, we analyze the spiking activity of a pyramidal cell in the CA1 region of the rat hippocampus recorded from an animal running back and forth on a linear track. Hippocampal pyramidal neurons have place-specific firing [29]. That is, a given neuron fires only when the animal is in a certain sub-region of the environment termed the neuron's place field. Because of this property these neurons are often called place cells. The neuron's spiking activity correlates most closely with the animal's position on the track ([38]). On a linear track these fields approximately resemble one-dimensional Gaussian functions. The data series we analyze consists of 4,265 spikes from a place cell in the CA1 region of the hippocampus recorded from a rat running back and forth for 1200 seconds on a 300-cm U-shaped track. In Figure 9.5, we show the linearized track plotted in time so that the spiking activity during the first 400 seconds can be visualized on a pass-by-pass basis [12]. We compare two approaches to estimating the place-specific firing maps of a hippocampal neuron. In the first, we use maximum likelihood to fit a specific parametric model of the spike times to the place cell data as in [3, 6, 8]. In the second approach we compute a histogram-based estimate of the conditional intensity function by using spatial smoothing of the spike train [12, 27]. The analysis presented here parallels the analysis performed in [8].

We let $x(t)$ denote the animal's position at time t, we define the spatial function for the one-dimensional place field model as the Gaussian function

$$s(t) = \exp\left\{\alpha - \frac{(x(t)-\mu)^2}{2\sigma^2}\right\}, \tag{9.34}$$

where μ is the center of the place field, σ^2 is a scale factor, and $\exp(\alpha)$ is the maximum height of the place field at its center. We define the spike time probability density of the neuron as either the inhomogeneous gamma (IG) model

$$p(u_j|u_{j-1},\theta) = \frac{\psi s(u_j)}{\Gamma(\psi)}\left[\int_{u_{j-1}}^{u_j}\psi s(u)du\right]^{\psi-1}\exp\left\{-\int_{u_{j-1}}^{u_j}\psi s(u)du\right\}, \tag{9.35}$$

or as the inhomogeneous inverse Gaussian (IIG) model

$$p(u_j|u_{j-1},\theta) = \frac{s(u_j)}{\left[2\pi\left[\int_{u_{j-1}}^{u_j}s(u)du\right]^3\right]^{1/2}}\exp\left\{-\frac{1}{2}\frac{\left(\int_{u_{j-1}}^{u_j}s(u)du-\psi\right)^2}{\psi^2\int_{u_{j-1}}^{u_j}s(u)du}\right\} \tag{9.36}$$

for $j = 1,\cdots,J$, where $\psi > 0$ is a location parameter for both models and $\theta = (\mu,\alpha,\sigma^2,\psi)$ is the set of model parameters to be estimated from the spike train. If we set $\psi = 1$ in Equation (9.35) we obtain the inhomogeneous Poisson (IP) model as a special case of the IG model. The models in Equations (9.35) and (9.36) are Markov because the current value of either the spike time probability density or the conditional intensity (rate) function (see Equation (9.3)) depends only on the time

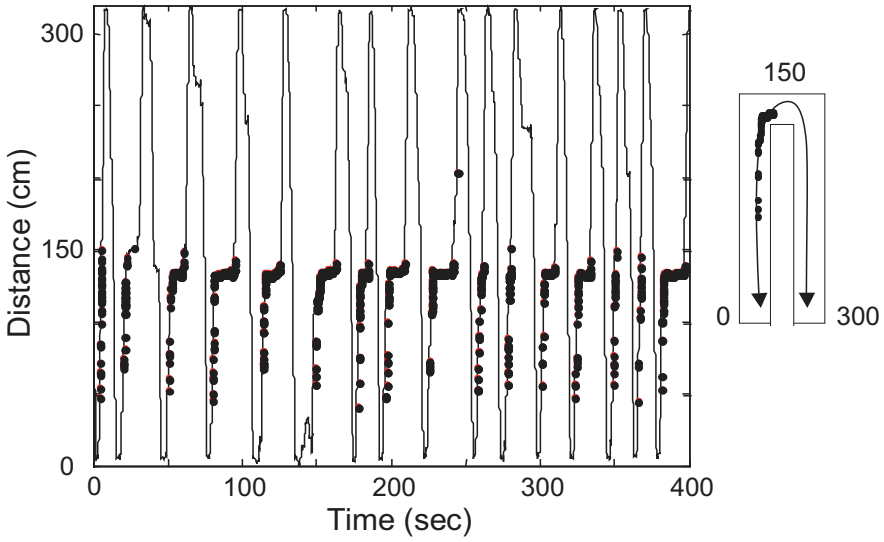

Figure 9.5

Place-specific spiking activity of a hippocampal pyramidal neuron recorded from a rat running back and forth for 400 sec of a 1200 sec experiment on a 300 cm U-shaped track (outset on the right). The track has been linearized and plotted in time so that the spiking activity on each pass can be visualized. The black dots show the spatial locations at which the neuron discharged a spike. The place field of this neuron extends from approximately 50 to 150 cms. In addition to having place-specific firing, this neuron is directional in that it spikes only as the animal moves from bottom to top (from left to right in the outset) between 50 to 150 cms.

of the previous spike. The IP, IG and IIG models are inhomogeneous analogs of the simple Poisson (exponential), gamma and inverse Gaussian models discussed in Section 9.3.1. These inhomogeneous models allow the spiking activity to depend on a temporal covariate, which, in this case, is the position of the animal as a function of time.

The parameters for all three models, the IP, IG and the IIG can be estimated from the spike train data by maximum likelihood [3, 8]. The log likelihoods for these two models have the form

$$\log L(\theta|N_{0:T}) = \sum_{j=1}^{J} \log p(u_j|u_{j-1},\theta) \tag{9.37}$$

To compute the spatial smoothing estimate of the conditional intensity function, we proceed as in [8]. We divide the 300 cm track into 4.2 cm bins, count the number of spikes per bin, and divide the count by the amount of time the animal spends in the bin. We smooth the binned firing rate with a six-point Gaussian window with a standard deviation of one bin to reduce the effect of running velocity [12]. The smoothed spatial rate function is the spatial conditional intensity estimate. The spatial smoothing procedure yields a histogram-based estimate of $\lambda(t)$ for a Poisson process because the estimated spatial function makes no history dependence assumption about the spike train. The IP, IG and IIG models were fit to the spike train data by maximum likelihood whereas the spatial rate model was computed as just described. As in Section 9.3.1 we use the KS plots to compare directly goodness-of-fit of the four models of this hippocampal place cell spike train.

The smoothed estimate of the spatial rate function and the spatial components of the rate functions for the IP, IG and IIG models are shown in Figure 9.6. The smoothed spatial rate function most closely resembles the spatial pattern of spiking seen in the raw data (Figure 9.5). While the spiking activity of the neuron is confined between approximately 50 and 150 cms, there is, on nearly each upward pass along the track, spiking activity between approximately 50 to 100 cms, a window of no spiking between 100 to 110 or 125 cms and then, a second set of more intense spiking activity between 125 to 150 cms. These data features are manifested as a bimodal structure in the smoothed estimate of the spatial rate function. The first mode is 10 spikes/sec and occurs at 70 cms, whereas the second mode is approximately 27 spikes/sec and occurs at approximately 120 cms. The spatial components of the IP and IG models were identical. Because of Equation (9.34), this estimate is unimodal and suggests a range of non-zero spiking activity which is slightly to the right of that estimated by the smoothed spatial rate function. The mode of the IP/IG model fits is 20.5 spikes/sec and occurs at approximately 110 cms. The IIG spatial component is also unimodal by virtue of Equation (9.34). It has its mode of 20.5 at 114 cms. This model significantly overestimates the width of the place field as it extends from 0 to 200 cm. The scale parameter, σ, is 23 cm for the IG model and 43 cm for the IIG model.

For only the IP and the smoothed spatial rate models do the curves in Figure 9.6 represent a spatial rate function. For the IG and IIG models the rate function, or

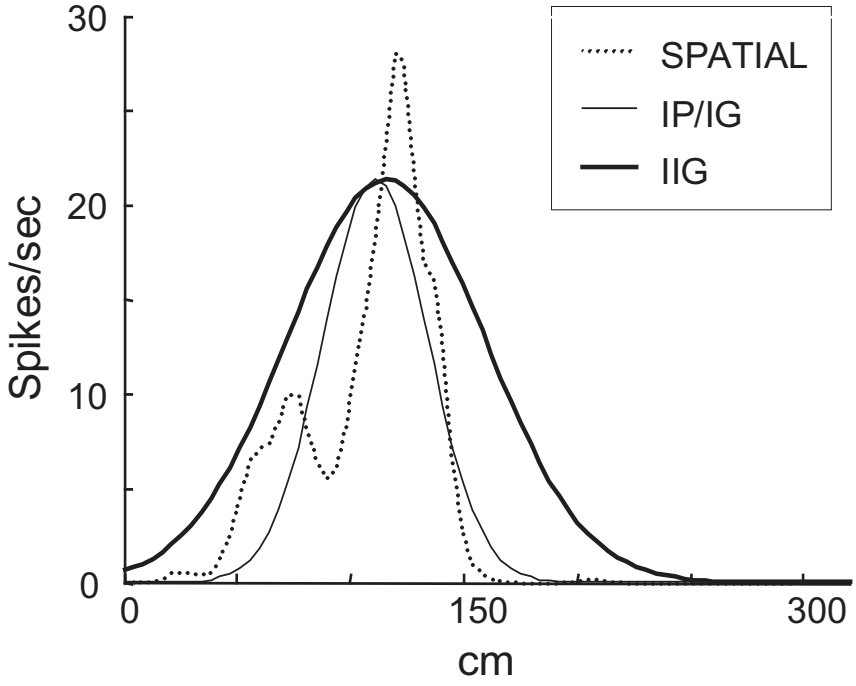

Figure 9.6

The place field estimates derived from the spatial smoothing model (dotted line), and the maximum likelihood fits of the inhomogeneous Poisson (IP) (thin solid line), inhomogeneous gamma (IG) (thin solid line), and inhomogeneous inverse Gaussian (IIG) models (thick solid line). The units of spikes/sec only apply to the spatial and IP model fits. For the IG and the IIG models the graphs show the spatial components of the rate models. The conditional intensity (rate) function for these two models is obtained from Equation (9.2) in the way Equation (9.33) was used to compute this quantity in Section 9.3.1.

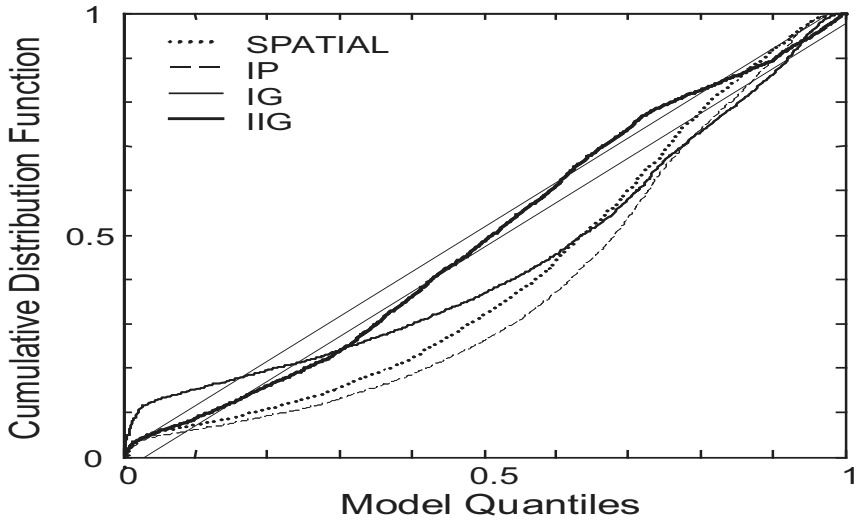

Figure 9.7

Kolmogorov-Smirnov plots of the spatial smoothing model (dotted line), and the maximum likelihood fits of the inhomogeneous Poisson (IP) (dashed line), inhomogeneous gamma (IG) (thin solid line), and inhomogeneous inverse Gaussian (IIG) (thick solid) models. As in Figure 9.4, the parallel diagonal lines are the 95% confidence bounds for the degree of agreement between the models and the spike train data.

conditional intensity function, is computed from Equation (9.2) using the maximum likelihood estimate of θ. For both of these models this equation defines a spatiotemporal rate function whose spatial component is defined by Equation (9.34). This is why the spatial components of these models are not the spatial rate function. For the IP model Equation (9.2) simplifies to Equation (9.34). The smoothed spatial rate model makes no assumption about temporal dependence, and therefore, it implicitly states that its estimated rate function is the rate function of a Poisson process.

The KS plot goodness-of-fit comparisons are shown in Figure 9.7. The IG model overestimates at lower quantiles, underestimates at intermediate quantiles, and lies within the 95% confidence bounds at the upper quantiles (Figure 9.7). The IP model underestimates the lower and intermediate quantiles, and like the IG model, lies within the 95% confidence bounds in the upper quantiles. The KS plot of the spatial rate model is similar to that of the IP model, yet closer to the confidence bounds. This analysis suggests that the IG, IP and spatial rate models are most likely oversmoothing this spike train because underestimating the lower quantiles corresponds to underestimating the occurrence of short ISIs [3]. The fact that the IP and IG mod-

els estimate a different temporal structure for this spike train data series is evidenced by the fact that while they both have the same spatial model components (Figure 9.6), Their KS plots differ significantly (Figure 9.7). This difference is due entirely to the fact that $\hat{\psi} = 0.61$ for the IG model whereas for the IP model $\psi = 1$ by assumption. The IP, the IG, and the smoothed spatial rate function have the greatest lack of fit in that order. Of the 4 models, the IIG is the one that is closest to the confidence bounds. Except for an interval around the 0.30 quantile where this model underestimates these quantiles, and a second interval around the 0.80 quantile where it overestimates these quantiles, the KS plot of the IIG model lies within the 95% confidence bounds.

The findings from the analysis of the spatial model fits (Figure 9.6) appear to contradict the findings of the overall goodness-of-fit analysis (Figure 9.7). The IIG gives the poorest description of the spatial structure in the data yet, the best overall description in terms of the KS plot. The smoothed spatial rate function model seems to give the best description of the data's spatial structure however, its overall fit is one of the poorest. To reconcile the findings, we first note that the better overall fit of the smoothed spatial rate function relative to the IP (Figure 9.7) is to be expected because the IP estimates the spatial component of a Poisson model with a three parameter model that must have a Gaussian shape. The smoothed spatial rate function model, on the other hand, uses a smoothed histogram that has many more parameters to estimate the spatial component of the same Poisson model. The greater flexibility of the smoothed spatial rate function allows it to estimate a bimodal structure in the rate function. Both the IP and smoothed spatial rate models use an elementary temporal model in that both assume that the temporal structure in the data is Poisson. For an inhomogeneous Poisson model the counts in non-overlapping intervals are independent whereas the interspike interval probability density is Markov (Equation (9.35)). The importance of correctly estimating the temporal structure is also seen in comparing the IP and IG model fits. These two models have identical spatial components yet, different KS plots because $\hat{\psi} = 0.61$ for the IG and $\psi = 1$ for the IP model by assumption. The KS plots suggest that while the IIG model does not describe the spatial component of the data well, its better overall fit comes because it does a better job at describing the temporal structure in the spike train. In contrast, the smoothed spatial rate function fits exclusively the spatial structure in the data to the exclusion of the temporal structure.

In summary, developing an accurate model of the place-specific firing activity of this hippocampal neuron requires specifying correctly both its spatial and temporal components. Our results suggest that combining a flexible spatial model, such as in the smoothed spatial rate function model with non-Poisson temporal structure as in the IG and IIG models, should be a way of developing a more accurate description. Another important consideration for hippocampal pyramidal neurons is that place-specific firing does not remain static. The current models would not capture this dynamic behavior in the data. In the next example we analyze a place cell from this same experiment using a point process adaptive filter algorithm to estimate the dynamics of the place cell spatial receptive fields using a model with flexible spatial and temporal structures.

9.3.3 An analysis of the spatial receptive field dynamics of a hippocampal neuron

The receptive fields of neurons are dynamic in that their responses to relevant stimuli change with experience. This plasticity, or experience-dependent change, has been established in a number of brain regions. In the rat hippocampus the spatial receptive fields of the CA1 pyramidal neurons evolve through time in a reliable manner as an animal executes a behavioral task. When, as in the previous example, the experimental environment is a linear track, these spatial receptive fields tend to migrate and skew in the direction opposite the cell's preferred direction of firing relative to the animal's movement, and increase in both maximum firing rate and scale [24, 25]. This evolution occurs even when the animal is familiar with the environment. As we suggested in Section 9.3.2, this dynamic behavior may contribute to the failure of the models considered there to describe the spike train data completely.

We have shown how the plasticity in neural receptive fields can be tracked on a millisecond time-scale using point process adaptive filter algorithms [7, 13]. Central to the derivation of those algorithms were the conditional intensity function (Equation (9.1)) and hte instantaneous log likelihood function (Equation (9.15)). We review briefly the derivation of the point process adaptive filter and illustrate its application by analyzing the spatial receptive field dynamics of a second pyramidal neuron from the linear track experiment discussed in Section 9.3.2.

To derive our adaptive point process filter algorithm we assume that the q-dimensional parameter θ in the instantaneous log likelihood (Equation (9.15)) is time varying. We choose K large, and divide $(0,T]$ into K intervals of equal width $\triangle = T/K$, so that there is at most one spike per interval. The adaptive parameter estimates will be updated at $k\triangle$. A standard prescription for constructing an adaptive filter algorithm to estimate a time-varying parameter is instantaneous steepest descent [17, 34] . Such an algorithm has the form

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \varepsilon \frac{\partial J_k(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}_{k-1}} \tag{9.38}$$

where $\hat{\theta}_k$ is the estimate at time $k\triangle$, $J_k(\theta)$ is the criterion function at $k\triangle$, and ε is a positive learning rate parameter. If for continuous-valued observations $J_k(\theta)$ is chosen to be a quadratic function of θ then, it may be viewed as the instantaneous log likelihood of a Gaussian process. By analogy, the instantaneous steepest descent algorithm for adaptively estimating a time-varying parameter from point process observations can be constructed by substituting the instantaneous log likelihood from Equation (9.15) for $J_k(\theta)$ in Equation (9.38). This gives

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \varepsilon \frac{\partial l_k(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}_{k-1}} \tag{9.39}$$

which, on rearranging terms, gives the instantaneous steepest descent adaptive filter algorithm for point process measurements

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \varepsilon \frac{\partial \log \lambda(k\triangle|H_k,\hat{\theta}_{k-1})}{\partial \theta}\left[dN(k\triangle) - \lambda(k\triangle|H_k,\hat{\theta}_{k-1})\triangle\right]. \tag{9.40}$$

Equation (9.40) shows that the conditional intensity function completely defines the instantaneous log likelihood and therefore, a point process adaptive filtering algorithm using instantaneous steepest descent. The parameter update $\hat{\theta}_k$ at $k\triangle$ is the previous parameter estimate $\hat{\theta}_{k-1}$ plus a dynamic gain coefficient,

$$-\frac{\varepsilon\partial \log \lambda(k\triangle|H_k,\hat{\theta}_{k-1})}{\partial\theta},$$

multiplied by an innovation or error signal $[dN(k\triangle)-\lambda(k\triangle|H_k,\hat{\theta}_{k-1})\triangle]$. The error signal provides the new information coming from the spike train and it is defined by comparing the predicted probability of a spike, $\lambda(k\triangle|\hat{\theta}_{k-1})\triangle$, at $k\triangle$ with $dN(k\triangle)$, which is 1 if a spike is observed in $((k-1)\triangle,k\triangle]$ and 0 otherwise. How much the new information is weighted depends on the magnitude of the dynamic gain coefficient. The parallel between the error signal in Equation (9.40) and that in standard recursive estimation algorithms suggests that the instantaneous log likelihood is a reasonable criterion function for adaptive estimation with point process observations. Other properties of this algorithm are discussed in [7].

Our objective is to identify plasticity related to both the spatial and temporal properties of the place receptive fields. Therefore, because given the learning rate, defining the conditional intensity function is sufficient to define the adaptive algorithm, we set

$$\lambda(k\triangle|H_k,\theta_k)=\lambda^S(k\triangle|x(k\triangle),\theta_k)\lambda^T(k\triangle-\zeta_k|\theta_k), \tag{9.41}$$

where $\lambda^S(k\triangle|x(k\triangle),\theta_k)$ is a function of the rat's position $x(k\triangle)$ at time $k\triangle$, $\lambda^T(k\triangle-\zeta_k|\theta_k)$ is a function of the time since the last spike, ζ_k is the time of the last spike prior to $k\triangle$ and θ_k is a set of time-dependent parameters. These two functions $\lambda^S(k\triangle|x(k\triangle),\theta_k)$ and $\lambda^T(k\triangle-\zeta_k|\theta_k)$ are respectively the spatial and temporal components of the conditional intensity function. To allow us to capture accurately the complex shape of place fields and the ISI structure of CA1 spike trains, we define $\lambda^S(k\triangle|x(k\triangle),\theta_k)$ and $\lambda^T(k\triangle-\zeta_k|\theta_k)$ as separate cardinal spline functions. A spline is a function constructed of piecewise continuously differentiable polynomials that interpolate between a small number of given coordinates, known as control point. The parameter vector θ_k contains the heights of the spline control points at time $k\triangle$. These heights are updated as the spikes are observed. As in [13] the spatial control points were spaced 1 every 10 cm plus one at each end for 32 total. The temporal control points were spaced 1 every 4 msec from 0 to 25 msec, and then every 25 msec from 25 to 1000 msec for 50 in total. Hence, the dimension of θ is $q=82$ in this analysis.

In Figure 9.8, we show the first 400 sec of spike train data from that experiment displayed with the track linearized and plotted in time as in Figure 9.5. The spiking activity of the neuron during the full 1200 sec of this experiment is used in the analysis. There were 1,573 spikes in all. The place-specific firing of the neuron is readily visible as the spiking activity occurs almost exclusively between 10 and 50 cms. As in the previous example, the spiking activity of the neuron is entirely unidirectional; the cell discharges only as the animal runs up and not down the track. The

intensity of spiking activity (number of spikes per pass) increases from the start of the experiment to the end.

We used the spline model of the conditional intensity function (Equation (9.41)) in the adaptive filter algorithm (Equation (9.40)) to estimate the dynamics of the receptive field of the neuron whose spiking activity is shown in Figure 9.8. The parameter updates were computed every 2 msec and the learning rate parameters were chosen based on the sensitivity analysis described in [13]. Examples of the spatial and temporal components of the conditional intensity function are shown in Figure 9.9. The migration of the spatial component during the course of the experiment is evidenced by the difference between these functions on the first pass compared with the last pass (Figure 9.9A). On the first pass the spatial function has a height of 12, is centered at approximately 40 cm and extends from 15 to 55 cms. By the last pass, the center of the spatial function has migrated to 52 cm, its height has increased to almost 20 and the range of the field extends from 15 to 70 cms. The migration of this spatial function is an exception. Typically, the direction of field migration is in the direction opposite the one in which the cell fires relative to the animal's motion [24, 25]. This place field migrates in the direction that the neuron fires relative to the animal's motion.

The temporal component of the intensity function characterizes history dependence as a function of the amount of time that has elapsed since the last spike. The temporal function shows increased values between 2 to 10 msec and around 100 msec. The former corresponds to the bursting activity of the neuron whereas the latter is the modulation of the place specific firing of the neuron by the theta rhythm [13]. For this neuron the modulation of the spiking activity due to the bursting activity is stronger than the modulation due to the approximately 6 to 14 Hz theta rhythm. Between the first and last pass the temporal component of the conditional intensity function increases slightly in the burst range and decreases slightly in the theta rhythm range. By definition, the rate function, i.e., the conditional intensity function based on the model in Equation (9.41) is the product of the spatial and temporal components at a given time. This is the reason why the units on the spatial and temporal components (Figure 9.9) are not spikes/sec. However, the product of the spatial and temporal components at a given time gives the rate function with units of spikes/sec. A similar issue arose in the interpretation of the spatial components of the conditional intensity functions for the IG and IIG models in Section 9.3.2 (Figure 9.6).

As in the previous two examples we used the KS plots based on the time-rescaling theorem to assess the goodness-of-fit of the adaptive point process filter estimate of the conditional intensity function (Figure 9.10). We compared the estimate of the conditional intensity function with and without the temporal component. The model without the temporal component is an implicit inhomogeneous Poisson process. The impact of including the temporal component is clear from the KS plots. For the model without the temporal component the KS plot does not lie within the 95% confidence bounds, whereas with the temporal component the plot is completely within the bounds. The improvement of the model fit with the temporal component is not surprising given that this component is capturing the effect of theta rhythm and

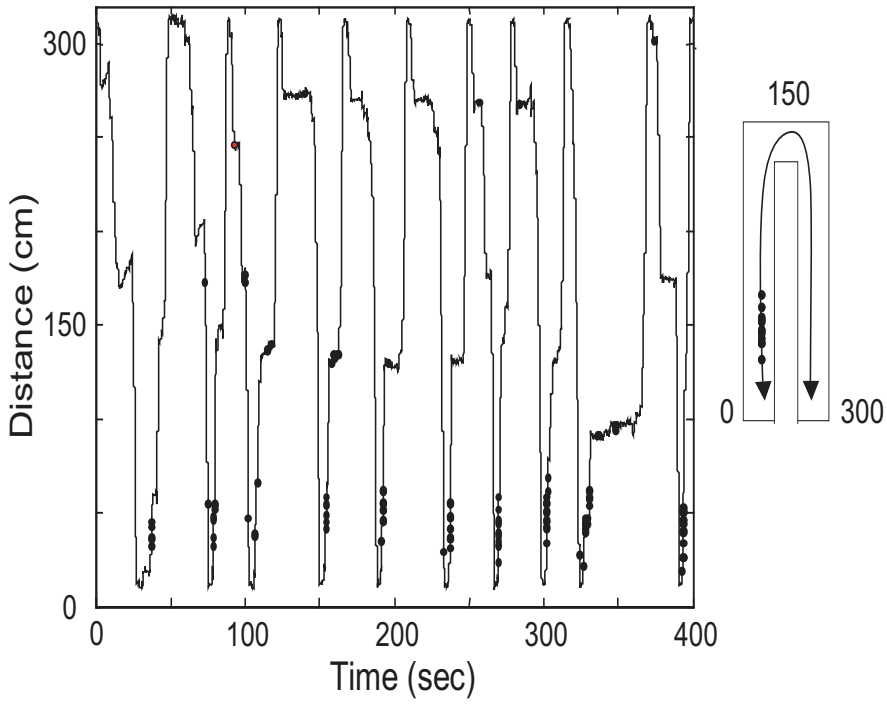

Figure 9.8

As in Figure 9.5, place-specific spiking activity of a second hippocampal pyramidal neuron recorded from a rat running back and forth for 400 sec of a 1200 sec experiment on a 300 cm U-shaped track (outset on the right). As in Figure 9.5, the track has been linearized and plotted in time so that the spiking activity on each pass can be visualized. The black dots show the spatial locations at which the neuron discharged a spike. The place field of this neuron extends from approximately 10 to 50 cms. Along with place-specific firing, this neuron is also directional in that it spikes only as the animal moves from bottom to top (from left to right in the outset) between 10 to 50 cms. The intensity of spiking increases from the start to the end of the experiment. In the data analyses we use the spiking activity during the entire 1200 sec.

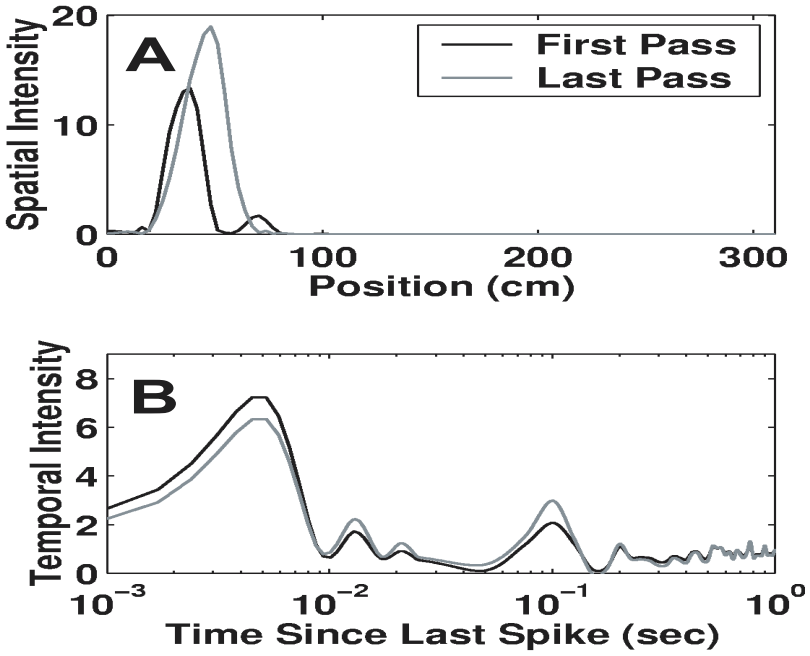

Figure 9.9

A. Point process adaptive filter estimates of the spatial component of the conditional intensity (rate) function on the first (solid black line) and last pass (solid gray line). B. Point process adaptive filter of the temporal component of the conditional intensity (rate) function on the first (solid black line) and last pass (solid gray line).

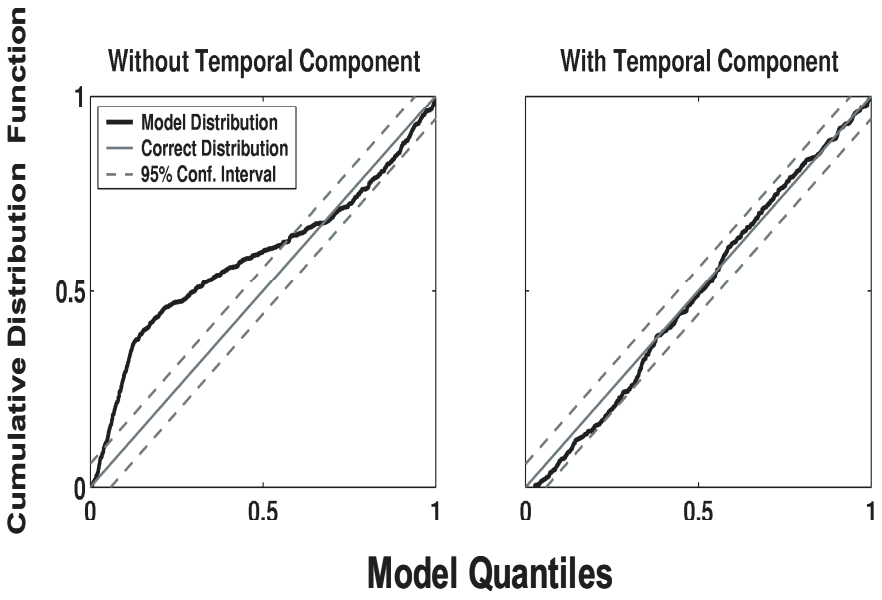

Figure 9.10

Kolmogorov-Smirnov (KS)plots for point process adaptive filter estimates of the conditional intensity (rate) function A. Without the temporal component of the model and B. With the temporal component of the model. The parallel dashed diagonal lines are the 95% confidence bounds for the degree of agreement between the models and the spike train data. The solid 45° line represents exact agreement between the model and the data. The adaptive filter estimate of the conditional intensity function with the temporal component gives a complete statistical description of this neural spike train based on the KS plot.

the bursting activity of this neuron (Figure 9.9B).

In summary, using the adaptive filter algorithm with the spline model to characterize the conditional intensity function we computed a dynamic estimate of the place receptive field. The updating was carried out on a 2 msec time-scale. We have found that the dynamics of these fields are best analyzed using videos. Videos of these types of analyses can be found on the websites cited in [7, 13]. These analyses show that use of a flexible model can lead to an accurate characterization of the spatial and temporal features of the hippocampal neuron's place receptive field. The results in this example illustrate an important improvement over the model fits in Section 9.3.2. These improvements can be measured through our KS plot goodness-of-fit tests. We believe these dynamic estimation algorithms may be used to characterize receptive field plasticity in other neural systems as well.

9.4 Conclusion

Neural spike trains are point processes and the conditional intensity function provides a canonical characterization of a point process. Therefore, we used the conditional intensity function to review several concepts and methods from likelihood theory for point process models that are useful for the analysis neural spike trains. By using the conditional intensity it was easy to show that the likelihood function of any point process model of a neural spike train has a canonical form given by Equation (9.8). The link between the conditional intensity function and the spike time probability model (Equation (9.2)) shows that defining one explicitly defines the other. This relation provided important modeling flexibility that we exploited in the analyses of three actual neural spike train data examples. In the first example, we used simple (renewal process) ISI models. In the second example, we applied spike time probability models that were modulated by a time-dependent covariate whereas in the third example, we formulated the spike train model directly in terms of the conditional intensity function. This allowed us to model explicitly history dependence and analyze the dynamic properties of the neurons receptive field. The conditional intensity function was also fundamental for constructing our goodness-of-fit tests based on the time-rescaling theorem.

The likelihood framework is an efficient way to extract information from a neural spike train typically by using parametric statistical models. We showed in the third example that it may also be used to develop dynamic estimation algorithms using a semiparametric model. Likelihood methods are some of the most widely used paradigms in statistical modeling due the extensive theoretical framework and the extensive applied experience that now lies behind these techniques.

Our analyses showed a range of ways of constructing and fitting non-Poisson models of neural spike train activity using the likelihood framework. While the different examples illustrated different features of the likelihood principles, we included in

each a goodness-of-fit analysis. We believe the goodness-of-fit assessment is a crucial, yet often overlooked, step in neuroscience data analyses. This assessment is essential for establishing what data features a model does and does not describe. Perhaps, most importantly, the goodness-of-fit analysis helps us understand at what point we may use the model to make an inference about the neural system being studied and how reliable that inference may be. We believe that greater use of the likelihood based approaches and goodness-of-fit measures can help improve the quality of neuroscience data analysis. Although we have focused here on analyses of single neural spike train time-series, the methods can be extended to analyses of multiple simultaneously recorded neural spike trains. These latter methods are immediately relevant as simultaneously recording multiple neural spike trains is now a common practice in many neurophysiological experiments.

9.5 Appendix

Lemma 1. Given n events $E_1, E_2, \cdots, E_n$ in a probability space, then

$$Pr\left(\cap_{i=1}^{n} E_i\right) = \prod_{i=2}^{n} Pr\left(E_i | \cap_{j=1}^{i-1} E_j\right) Pr(E_1). \tag{9.42}$$

Proof: By the definition of conditional probability for $n = 2$, $Pr(E_1 \cap E_2) = Pr(E_2|E_1)Pr(E_1)$. By induction

$$Pr\left(\cap_{i=1}^{n-1} E_i\right) = \prod_{i=2}^{n-1} Pr\left(E_i | \cap_{j=1}^{i-1} E_j\right) Pr(E_1). \tag{9.43}$$

Then

$$\begin{aligned} Pr\left(\cap_{i=1}^{n} E_i\right) &= Pr\left(E_n | \cap_{i=1}^{n-1} E_i\right) Pr\left(\cap_{i=1}^{n-1} E_i\right) \\ &= Pr\left(E_n | \cap_{i=1}^{n-1} E_i\right) \prod_{i=1}^{n-1} Pr\left(E_i | \cap_{j=1}^{i-1} E_j\right) Pr(E_1) \\ &= \prod_{i=1}^{n} Pr\left(E_i | \cap_{j=1}^{i-1} E_j\right) Pr(E_1). \qquad \textbf{Q.E.D.} \end{aligned} \tag{9.44}$$

Acknowledgments. Support for this work was provided in part by NIMH grants MH59733, MH61637, NIDA grant DA015664 and NSF grant IBN-0081458. We are grateful to Satish Iyengar for supplying the retinal spike train data analyzed in Section 9.3.1.

References

[1] Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models based on Counting Processes*. New York: Springer-Verlag.

[2] Barbieri, R., Frank, L. M. Quick, M. C., Wilson, M. A., and Brown, E. N. (2001) Diagnostic methods for statistical models of place cell spiking activity. *Neurocomputing*, **38-40**, 1087-1093.

[3] Barbieri, R., Quirk, M.C., Frank, L. M., Wilson, M. A., and Brown, E. N. (2001). Construction and analysis on non-Poisson stimulus-response models of neural spike train activity. *J. Neurosci. Meth.*, **105**, 25-37.

[4] Berman, M. (1983). Comment on "Likelihood analysis of point processes and its applications to seismological data" by Ogata. *Bulletin Internatl. Stat. Instit.*, **50**, 412-418.

[5] Brillinger, D. R. (1988). Maximum likelihood analysis of spike trains of interacting nerve cells. *Biol. Cyber.*, **59**, 189-200.

[6] Brown, E. N., Frank, L. M., Tang, D., Quirk, M. C., and Wilson, M. A. (1998). A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, **18**, 8411-7425.

[7] Brown, E. N., Nguyen, D. P., Frank, L. M., Wilson, M. A., and Solo V. (2001). An analysis of neural receptive field plasticity by point process adaptive filtering. *PNAS*, **98**, 12261-12266.

[8] Brown, E. N., Barbieri, R., Ventura, V., Kass, R. E., and Frank, L. M. (2002). The time-rescaling theorem and its application to neural spike train data analysis. *Neural Comput.*, **14**, 325-346.

[9] Casella, G., and Berger, R. L. (1990). *Statistical Inference*. Belmont, CA: Duxbury.

[10] Chhikara, R. S., and Folks, J. L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications.* New York: Marcel Dekker.

[11] Daley, D., and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* New York: Springer-Verlag.

[12] Frank, L. M., Brown, E. N., and Wilson, M. A., (2000). Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, **27**, 169-178.

[13] Frank, L. M., Eden U. T., Solo, V., Wilson, M. A., and Brown, E. N., (2002). Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. *Journal of Neuroscience*, **22**, 3817-3830.

[14] Gerstein, G. L. and Mandelbrot, B. (1964) Random walk models for the spike activity of a single neuron. *J. Biophys.*, **4**, 41-68.

[15] Gerstner, W., and Kistler, W. M. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge, UK: University Press.

[16] Guttorp, P. (1995). *Stochastic Modeling of Scientific Data*. London: Chapman and Hall.

[17] Haykin, S. (1996). *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall.

[18] Iyengar, S., and Liao, Q. (1997). Modeling neural activity using the generalized inverse Gaussian distribution. *Biol. Cyber.*, **77**, 289-295.

[19] Johnson, A., and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions-2*. New York: Wiley.

[20] Kalbfleisch, J., and Prentice, R. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

[21] Kass, R. E., and Ventura, V. (2001). A spike train probability model. *Neural Comput.*, **13**, 1713-1720.

[22] Levine, M. W., Saleh, E. J., and Yamold, P. (1988). Statistical properties of the maintained discharge of chemically isolated ganglion cells in goldfish retina. *Vis. Neurosci.*, **1**, 31-46.

[23] Levine, M. W. (1991). The distribution of intervals between neural impulses in the maintained discharges of retinal ganglion cells. *Biol. Cybern.*, **65**, 459-467.

[24] Mehta, M. R., Barnes, C. A., and McNaughton, B. L. (1997). *Proc. Natl. Acad. Sci. USA*, **94**, 8918-8921.

[25] Mehta, M. R., Quirk, M. C., and Wilson, M. A. (2000). *Neuron*, **25**, 707- 715.

[26] Meyer, P. (1969). Démonstration simplifiée d'un théoréme de Knight. In *Séminaire Probabilité V* (pp. 191-195). New York: Springer-Verlag.

[27] Muller, R. U., and Kubie, J. L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, **7**, 1951-1968.

[28] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of American Statistical Association*, **83**, 9-27.

[29] O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.*, **34**, 171-175.

[30] Papangelou, F. (1972). Integrability of expected increments of point processes and a related random change of scale. *Trans. Amer. Math. Soc.*, **165**, 483-506.

[31] Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood.* London: Oxford.

[32] Schroedinger, E. (1915). Zur Theorie der fall- und steigversuche an teilchen mit Brownscher bewegung. *Phys. Ze.*, **16**, 289-295.

[33] Snyder, D., and Miller, M. (1991). *Random Point Processes in Time and Space* (2nd ed.). New York: Springer-Verlag.

[34] Solo, V. and Kong, X. (1995). *Adaptive Signal Processing Algorithms: Stability and Performance.* Upper Saddle River, NJ: Prentice-Hall.

[35] Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* In Springer Series in Statistics, New York: Springer-Verlag.

[36] Taylor, H. M., and Karlin, S. (1994). *An Introduction to Stochastic Modeling* (rev. ed.) San Diego, CA: Academic Press.

[37] Tuckwell, H. C. (1988). *Introduction to Theoretical Neurobiology: Nonlinear and Stochastic Theories.* New York; Cambridge.

[38] Wilson, M. A., and McNaughton, B. L., (1993). Dynamics of the hippocampal ensemble code for space. *Science*, **261**, 1055-1058.

[39] Wood, E. R., Dudchenko, P. A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, **397**, 613- 616.