SUMMER SCHOOL IN THEORETICAL PHYSICS
LES HOUCHES, FRANCE

THEORY OF POINT PROCESSES FOR NEURAL SYSTEMS

EMERY N. BROWN

DEPARTMENT OF ANESTHESIA AND CRITICAL CARE
MASSACHUSETTS GENERAL HOSPITAL

DIVISION OF HEALTH SCIENCES AND TECHNOLOGY
HARVARD MEDICAL SCHOOL
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

AUGUST 4, 2003

Acknowledgements: Support for this work was provided in part by NIH grants MH66410, MH59733, MH61637 and DA015664. We thank Uri Eden and Riccardo Barbieri for help preparing the figures.

1. Neural spike trains as point processes

Modeling analyses of neural systems are typically performed with Hodgkin and Huxley, integrate-and-fire and neural network models. In general, these models treat the processes of action potential production as deterministic. Much insight in the behavior of neural systems has been obtained from these kinds of modeling analyses. However, for actual neurons, the deterministic representation is never completely true as many factors which these models assume are rarely known with certainty, even in controlled experiments. Indeed, the non deterministic nature of neural processes is readily apparent from the plot of any neural spike train recorded in a neurophysiological study. In general, the deterministic models cannot suggest strategies or methods to analyze the non-deterministic properties of neural spike trains. Therefore, it is important to have a stochastic framework in which to model neural processes and to analyze neural spike train data.

While action potentials are not instantaneous, it is typical to assign them occurrence times such as the time at which the membrane voltage crosses threshold. The discrete, all-or- nothing nature of a sequence of neuronal action potentials together with their stochastic structure suggests that a neuronal spike train can be viewed as a point process. A point process is a stochastic process composed of a sequence of binary events that occur in continuous time. The theory of point processes is a highly developed subdiscipline in the field of stochastic processes [1]. There has recently been extensive theoretical study of point processes as well as application of this theory in biostatistics, geophysics and stochastic control. In many cases, these results have been presented in a highly theoretical framework that is not accessible to most students of computational neuroscience. Often the more applied work is not related to the types of point process models that are relevant for neural systems.

These notes offer an introduction to the theory of univariate point processes for neural systems. We focus on results from the theory of univariate point processes that are relevant for modeling and data analysis in neuroscience. Section 2 reviews the derivation of interspike interval models from elementary stochastic dynamical systems models. Section 3 defines the conditional intensity function for a point process and discusses its properties. Section 4 derives the joint probability density of a point process. Section 5 discusses some special point process models including the Poisson process, renewal processes and stationary processes. Section 6 derives the time-rescaling theorem and discusses its implications for neural spike train data analysis. Section 7 discusses methods for simulating point processes. Section 8 mentions briefly Poisson limit results.

2. Integrate and fire models and interspike interval distributions

There are two primary ways to characterize a point process. The first is in terms of the interevent probability model and the second is in terms of the conditional intensity function. The intervent probability model is the interspike interval probability model whereas the conditional intensity function, in its most general sense, is a history dependent rate function. Defining one defines the other and vice-versa. It is important to develop facility with using the two characterizations and to appreciate the interrelation between them. In this section, we derive the relation between elementary integrate-and-fire neuron models and the interspike interval distributions. We begin our discussion with these elementary stochastic models. This will make explicit the relation between the stochastic versions of elementary dynamical systems, neural models and statistical models used to analysis of neural data. In Section 3, we develop the characterization of a point process in terms of its conditional intensity function and we relate it to the interspike interval distribution. This section follows closely [2].

2.1. Non-leaky integrator with excitatory Poisson inputs

Consider a neuron whose membrane voltage time course is defined by

$$dV(t) = \alpha_E dN(t), \tag{2.1}$$

where $N(t)$ is a Poisson process with constant rate parameter λ, and α_E is the magnitude of each excitatory input (**Fig. 1A**). The solution to Eq. 2.1 is

$$V(t) = \int_0^t \alpha_E dN(u) = \alpha_E N(t). \tag{2.2}$$

Suppose that the resting membrane potential at time 0 is $V(0) = 0$ and the neuron discharges an action potential when $V(t) \geq \theta$, where θ is a constant threshold voltage. Notice that for $V(t) \geq 0$ we must have $\alpha_E N(t) \geq \theta$ or $N(t) \geq \theta \alpha_E^{-1}$. If we let $[x]$ denote the greatest integer $\leq x$ then we require $N(t) > 1 + [\theta \alpha_E^{-1}],$ to observe an action potential.

Figure 1 about here.

What is the probability density of the times between the action potentials? To evaluate this probability we need to compute the probability density of the waiting time until the k^{th} event for a Poisson process, beginning at an arbitrary time point s. This is

$$\begin{aligned}
p_k(t)\Delta &= \Pr(k^{th}\ \ event\ \ in(s+t, s+t+\Delta]) \\
&= \Pr(k-1st\ event\ in(s, s+t] \bigcap 1 event(s+t, s+t+\Delta]) \\
&= \Pr(k-1st\ event\ in(s, s+t]) \Pr(1\ event(s+t, s+t+\Delta] \mid k-1st\ \ event\ in(s, s+t]) \\
&= \Pr(k-1st\ event\ in(s, s+t]) \Pr(1\ event(s+t, s+t+\Delta])
\end{aligned}$$

$$= \frac{e^{-\lambda t}(\lambda t)^{k-1}}{(k-1)!} .\lambda\Delta . \tag{2.3}$$

Hence,

$$p_k(t)\Delta = \frac{e^{-\lambda t}(\lambda\Delta)^{k-1}}{(k-1)!}\lambda\Delta,$$

which gives

$$p_k(t) = \frac{e^{-\lambda t}(\lambda t)^{k}}{(k-1)!} = \frac{e^{-\lambda t}\lambda t^{(k-1)-1}}{\Gamma(k)}. \tag{2.4}$$

We see that $p_k(t)$ is a gamma probability density with parameters k and λ. We have $E(t) = k\lambda^{-1}$ and $Var(t) = k\lambda^{-2}$ (**Fig. 2**).

For the primitive neuron model, we need $k = 1+[\theta\alpha_E^{-1}]$ to generate a spike. Hence, the interspike interval probability density is the gamma probability density with parameters $1+\theta\alpha_E^{-1}$ and λ. The interspike probability density is exponential if and only if $k = 1$. Hence, Poisson models are not the point process model associated with even elementary spike train models. The shape of the probability density is right-skewed.

Figure 2 about here.

2.2. Non-leaky integrator with excitatory and inhibitory Poisson inputs

We next consider a non-leaky integrator neuron with both excitatory and inhibitory Poisson inputs. For this neuron the time course of the membrane voltage is

$$dV(t) = \alpha_E dN_E(t) - \alpha_I dN_I(t), \tag{2.5}$$

where $N_E(t) \sim P(\lambda_E)$ and $N_I(t) \sim P(\lambda_I)$ are independent Poisson processes that govern the respective times of the excitatory and inhibitory inputs and, α_E and α_I are respectively the magnitudes of the inhibitory and excitatory inputs. We have

$$V(t) = \alpha_E\int_0^t dN_E(u) - \alpha_I\int_0^t dN_I(u) = \alpha_E N(t) - \alpha_I N_I(t) . \tag{2.6}$$

For this model what is the interspike interval probability density? The interspike interval probability density for this model when $\alpha_E = \alpha_I = 1$ is given by [2]

$$p_\theta(t) = \theta\left(\frac{\lambda_E}{\lambda_I}\right)^{\theta/2}\frac{e^{-(\lambda_E+\lambda_I)t}}{t} I_\theta\left(2t(\lambda_E\lambda_I)^{\frac{1}{2}}\right) \quad t > 0 \tag{2.7}$$

for $t > 0$ where

$$I_\rho(x) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(k+\rho+1)} \left(\frac{x}{2}\right)^{2k+\rho},$$

is a modified Bessel function. We also have that

$$\Pr(t_\theta < \infty) = \begin{cases} 1 & \lambda_I \le \lambda_E \\ {\lambda_E}^{\theta} {\lambda_I}^{-1} & \lambda_E < \lambda_I. \end{cases} \tag{2.8}$$

Equation 2.8 means that if $\lambda_E \ge \lambda_I$ the time to threshold is finite with probability 1 whereas if $\lambda_E < \lambda_I$ threshold may not be reached. This model is analytically tractable and leads to the Wiener process model we discuss next as the limiting case.

2.3. Non-leaky integrator with random walk inputs

Finally, we consider a non-leaky integrator neuron with Gaussian random walk inputs. That is, we define the membrane voltage equation as (**Fig. 1B**)

$$dV_\alpha(t) = \alpha dN_E(t) - \alpha dN_I(t), \tag{2.9}$$

or

$$V_\alpha(t) = \alpha \int_0^t dN_E(u) - \int_0^t dN_I(u) = \alpha[N_E(t) - N_I(t)], \tag{2.10}$$

where $\alpha_E = \alpha_I = \alpha$ and $\lambda_E = \lambda_I$ and $E[V(t)] = 0$ and $Var[V(t)] = 2\alpha^2 \lambda t.$ We have as $V_\alpha(t) \to W(t),$ i.e., $V_\alpha(t)$ converges to $W(t)$ in distribution, where $W(t)$ is a Wiener process. The Wiener process (Brownian motion), $W(t), t \ge 0$, is defined by the following three properties

i. $W(0) = 0$.
ii. If $(t_j, t_{j+1}]$ and $(t_k, t_{k+1}]$ are non overlapping intervals, then $W(t_{j+1}) - W(t_j)$ and $W(t_{k-1}) - W(t_k)$ are independent.
iii. $W(t_{k+1}) - W(t_k) \sim N(0, \sigma^2(t_{k+1} - t_k))$.

Its probability density function is given as

$$p_w(w \mid t) = (2\pi t)^{-\frac{1}{2}} \exp\{-(2t)^{-1} w^2\}. \tag{2.11}$$

From this model we can define the Wiener process with drift by making the transformation

$$V(t) = V_0 + \sigma W(t) + \beta t, \tag{2.12}$$

and we find that this is a Gaussian process with mean and variance defined as

$$E[V(t)] = V_0 + \beta t \tag{2.13}$$

$$Var[V(t)] = \sigma^2 t\,. \tag{2.14}$$

If the primitive neuron receives stochastic inputs for a Wiener process or a Wiener process with drift then, what is the interspike interval probability density? We define the first passage time as the condition $t_\theta = \inf\{u \mid V(u) = \theta\}$, $V(0) = V_0 < \theta$. We consider two cases [2,3]. First we start with the driftless Wiener process, that is with $\beta = 0$. For the driftless Wiener process model the first passage time probability density is

$$p_\theta(t) = \frac{\theta - V_0}{(2\pi\sigma^2 t^3)^{\frac{1}{2}}} \exp\left\{-\frac{(\theta - V_0)^2}{2\sigma^2 t}\right\}, \tag{2.15}$$

$t > 0 \;\; \theta > x_0$ (**Fig. 3**). For the Wiener process model with drift the first passage time probability density is given by the inverse Gaussian probability density[3,4] defined as

$$p_\theta(t) = \frac{\theta - V_0}{(2\pi\sigma^2 t^3)^{\frac{1}{2}}} \exp\left\{\frac{(\theta - V_0 - \beta t)^2}{2\sigma^2 t}\right\}, \tag{2.16}$$

$t > 0 \;\; \theta > x_0$. The probability of reaching threshold in finite time is defined by

$$\Pr\{t_\theta < \infty\} = \begin{cases} 1 & \beta \geq 0 \\ \exp\left[-\dfrac{2\,|\,\beta\,|\,(\theta - V_0)}{\sigma^2}\right], & \beta < 0. \end{cases} \tag{2.17}$$

For the inverse Gaussian probability density, the mean and variance are

$$E[t_\theta] = \frac{(\theta - V_0)}{\beta} = \frac{\theta}{\alpha_E \lambda_E - \alpha_{I\lambda_I}} \qquad V[t_\theta] = \frac{(\theta - V_0)\sigma^2}{\beta^3} = \frac{\theta\alpha_E^2 \lambda_E + \alpha_I^2 \lambda_I}{(\alpha_E \lambda_E - \alpha_I \lambda_I)^3}\,, \tag{2.18}$$

$\beta > 0,\, \theta \geq x_0$. The coefficient of variation is

$$CV = \frac{Var[t_\theta]^{\frac{1}{2}}}{E[t_\theta]} = \left[\frac{\alpha_E \lambda_E + \alpha_I \lambda_I}{\alpha_E \lambda_E - \alpha_{I\lambda_I}}\right]^{\frac{1}{2}}. \tag{2.19}$$

Figure 3 about here.

2.4. Remarks

1. If we let the parameter be defined as $\mu = E[t_\theta]$ and $V[t_\theta] = \mu^3 \lambda^{-1}$ then, under this change of parameters the inverse Gaussian probability density can be expressed as

$$p(t|\mu,\lambda)=\left(\frac{\lambda}{2\pi t^3}\right)^{\frac{1}{2}}\exp\left\{-\frac{1}{2}\frac{\lambda(t-\mu)^2}{\mu^2 t}\right\}. \tag{2.20}$$

2. Schrödinger in (1915) gave one of the first derivations of this probability model [3].

3. The probability density is right-skewed.

4. If $(\mu/\lambda)^{-1}$ is large then, the inverse Gaussian probability density is well approximated by a Gaussian probability density.

5. If we record only the spikes then, we are only able to recover the two parameters μ and λ whereas in the original random walk model, there are four parameters V_0, β, θ and σ^2.

6. The results from this section show that if we construct even the simplest of dynamical models of neurons then, the resulting probability model for the interspike interval distribution are all right-skewed and non-exponential (**Figs 2, 3**). Hence, exponential waiting times or Poisson models are not the elementary models most likely to agree with neural spike train data.

3. The conditional intensity function and interevent time probability density

Neural spike trains are characterized by their interspike interval probability models. In Section 2 we showed how elementary interspike interval probability models can be derived from elementary stochastic dynamical systems models of a neuron. By viewing the neural spike trains as a point process we can present another characterization of the spike train in terms of its conditional intensity function. We develop this characterization in this section and we relate the conditional intensity function to the interspike interval probability models in Section 2. The presentation here follows closely [5].

Let $(0,T]$ denote the observation interval and let $0<u_1<u_2<,\ldots,<u_{J-1}<u_J\le T$ 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 $N_{0:t}=\{0<u_1<u_2,\ldots,u_j\le t\cap N(t)=j\}$, where $N(t)$ is the number of spikes in $(0,t]$ and $j\le J$. The sample path is a right continuous function that jumps 1 at the spike times and is constant otherwise [1,5,6,7,8]. 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 (**Fig. 3A**). The counting process $N(t)$ gives the total number of events that have occurred up to time *t*. The counting process satisfies

i) $N(t)\ge 0.$

ii) $N(t)$ is an integer-valued function.

iii) If $s<t,$ then $N(s)\le N(t)$.

iv) For $s<t,\ N(t)-N(s)$ is the number of events in (s,t).

Figure 4 about here.

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

$$\lambda(t \mid H_t) = \lim_{\Delta \to 0} \frac{\Pr(N(t+\Delta) - N(t) = 1 \mid H_t)}{\Delta}, \tag{3.1}$$

where H_t is the history of the sample path and of any covariates up to time t. In general $\lambda(t \mid H_t)$ depends on the history of the spike train and therefore, it is also termed the stochastic intensity. In survival analysis the conditional intensity function is called the hazard function. [9,10] This is because the hazard function can be used to define the probability of an event in the interval $[t,t+\Delta)$ given that there has not been an event up to t. For example it might represent the probability that a piece of equipment fails in $[t,t+\Delta)$ given that it was worked up to time t.[9] As another example, it might define the probability that a patient receiving a new therapy dies in the interval $[t,t+\Delta)$ given that he/she has survived up to time t.[10] It follows that $\lambda(t \mid H_t)$ can be defined in terms of the interspike interval probability density at time t, $p(t \mid H_t)$, as

$$\lambda(t \mid H_t) = \frac{p(t \mid H_t)}{1 - \int_0^t p(u \mid H_t)du} \tag{3.2}$$

We gain insight into the definition of the conditional intensity function in Eq. 3.1 by considering the following heuristic derivation of Eq. 3.2 based on the definition of the hazard function. We compute explicitly the probability of the event, a spike in $[t,t+\Delta)$ given H_t and that there has been no spike in $(0,t)$. That is,

$$\Pr(u \in [t,t+\Delta) \mid u > t, H_t) = \frac{\Pr(u \in [t,t+\Delta) \cap u > t \mid H_t)}{\Pr(u > t \mid H_t)}$$

$$= \frac{\Pr(u \in [t,t+\Delta) \mid H_t)}{\Pr(u > t \mid H_t)}$$

$$= \frac{\int_t^{t+\Delta} p(u \mid H_u)du}{1 - \int_0^t p(u \mid H_u)du}$$

$$= \frac{p(t \mid H_t)\Delta}{1 - \int_o^t p(u \mid H_u)du} + o(\Delta)$$

$$= \lambda(t \mid H_t)\Delta + o(\Delta) \tag{3.3}$$

where $o(\Delta)$ refers to all events of order smaller than Δ, such as two or more spikes occurring in an arbitrarily small interval. This establishes Eq. 3.2.

The power of the conditional intensity function is that if it can be defined, as Eq 3.3 suggests then, it completely characterizes the stochastic structure of the spike train. In any time interval $[t,t+\Delta)$, $\lambda(t \mid H_t)\Delta$ defines the probability of a spike given the history up to time t. If the spike train is an inhomogeneous Poisson process then, $\lambda(t \mid H_t) = \lambda(t)$ becomes the Poisson rate function. Thus, the conditional intensity function (Eq. 3.1) is a history-dependent rate function that generalizes the definition of the Poisson rate. Similarly, Eq. 3.1 is also a generalization of the hazard function for renewal processes [9,10].

Example 3.1. Conditional Intensity Function of the Gamma Probability Density. The gamma probability density for the integrate and fire model in Eq. 2.4 is

$$p_k(t) = \frac{e^{-\lambda t}\lambda t^{(k-1)-1}}{\Gamma(k)}. \tag{3.4}$$

From Eq. 3.2, it follows that the conditional intensity function is

$$\lambda(t) = \frac{e^{-\lambda t}\lambda t^{(k-1)-1}}{\Gamma(k)\left[1-\int_0^t p_k(u)du\right]}. \tag{3.5}$$

Example 3.2. Conditional Intensity Function of the Inverse Gaussian Probability Density. The inverse Gaussian probability density for the Wiener process integrate and fire model in Eq. 2.20 is

$$p(t|\mu,\rho) = \left(\frac{\rho}{2\pi t^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{\rho(t-\mu)^2}{\mu^2 t}\right\}. \tag{3.6}$$

From Eq. 3.2 the conditional intensity function for this model is

$$\lambda(t|\mu,\rho) = \frac{\left(\frac{\rho}{2\pi t^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{\rho(t-\mu)^2}{\mu^2 t}\right\}}{1-\int_0^t\left(\frac{\rho}{2\pi u^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\frac{\rho(u-\mu)^2}{\mu^2 u}\right\}du}. \tag{3.7}$$

4. Joint probability density of a point process [5]

The likelihood function is a primary tool used in constructing statistical models [5]. The likelihood function 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 and hence, spike train is easy to derive from the conditional intensity function. We begin with a lemma.

4.1. Derivation of the joint probability density

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

$$\Pr(\bigcap_{i=1}^{n} E_i) = \prod_{i=1}^{n} \Pr(E_i \mid \bigcap_{j=1}^{i-1} E_j) \Pr(E_1). \tag{4.1}$$

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

$$\Pr(\bigcap_{i=1}^{n-1} E_i) = \prod_{2=1}^{n-1} \Pr(E_i \mid \bigcap_{j=1}^{i-1} E_j) \Pr(E_1). \tag{4.2}$$

Then

$$\begin{aligned}
\Pr(\bigcap_{i=1}^{n} E_i) &= \Pr(E_n \mid \bigcap_{i=1}^{n-1} E_i) \Pr(\bigcap_{i=1}^{n-1} E_i) \\
&= \Pr(E_n \mid \bigcap_{i=1}^{n-1} E_i) \prod_{i=1}^{n-1} \Pr(E_i \mid \bigcap_{j=1}^{i-1} E_j) \Pr(E_1) \qquad (4.3) \\
&= \prod_{i=1}^{n} \Pr(E_i \mid \bigcap_{j=1}^{i-1} E_j) \Pr(E_1).
\end{aligned}$$

Q.E.D.

.

Theorem 1. Given $0 < u_1, < u_2, \ldots, u_J < T$, 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

$$\begin{aligned}
p(N_{0:T}) &= \prod_{j=1}^{J} \lambda(u_j \mid H_{u_j}) \exp\left\{-\int_0^T \lambda(u \mid H_u) du\right\} \\
&= \exp\left\{\int_0^T \log \lambda(u \mid H_u) dN(u) - \int_0^T \lambda(u \mid H_u) du\right\}. \qquad (4.4)
\end{aligned}$$

Proof. Let $\{t_k\}_{k=1}^{K}$ be a partition of the observation interval $(0,T]$. Take $\Delta_k = t_k - t_{k-1}$, 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 $\Delta_k \leq 1$ 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

$$A_k = \{\text{spike in } (t_{k-1}, t_k] \mid H_k\}$$

$$E_k = \{A_k\}^{dN(k)}\{A_k^c\}^{1-dN(k)} \tag{4.5}$$

$$H_k = \left\{\bigcap_{j=1}^{k-1} E_j\right\},$$

for $k = 1, \ldots, K$. In any interval $(t_{k-1}, t_k]$ we have (**Fig. 3B**)

$$\Pr(E_k) = \begin{cases} \lambda(t_k \mid H_k)\Delta_k + o(\Delta_k) & \text{if } dN(k) = 1 \\ 1 - \lambda(t_k \mid H_k)\Delta_k + o(\Delta_k) & \text{if } dN(k) = 0 \end{cases} \tag{4.6}$$

By construction of the partition we must have $u_j \in (t_{k_j-1}, t_{k_j}]$, $j = 1, \ldots, J$ for a subset of the intervals satisfying $k_1 < k_2 < \ldots < k_J$. The remaining $K - J$ intervals have no spikes. The spike events form a sequence of correlated Bernoulli trials. It follows from Eq. 4.6 and **Lemma 1** that given the partition, the probability of exactly J events in $(0, T]$ may be computed as

$$p(N_{0:T})\prod_{j=1}^{J}\Delta_{k_j} = p(u_j \in (t_{k_j-1}, t_{k_j}], j = 1, \ldots, J \cap N(T) = J)\prod_{j=1}^{J}\Delta_{k_j}$$

$$= \Pr\left(\bigcap_{k=1}^{K} E_k\right)$$

$$= \prod_{k=2}^{K}\Pr\left(E_k \mid \bigcap_{j=1}^{k-1} E_j\right)\Pr(E_1)$$

$$= \prod_{k=1}^{K}[\lambda(t_k \mid H_k)\Delta_k]^{dN(t_k)}[1 - \lambda(t_k \mid H_k)\Delta_k]^{1-dN(t_k)} + o(\Delta_*)$$

$$= \prod_{j=1}^{J}[\lambda(t_{k_j} \mid H_{k_j})\Delta_{k_j}]^{dN(t_{k_j})}\prod_{\ell \neq k_j}[1 - \lambda(t_\ell \mid H_\ell)\Delta_\ell]^{1-dN(t_\ell)}$$

$$+ o(\Delta_*)$$

$$= \prod_{j=1}^{J}\lambda(t_{k_j} \mid H_{k_j})\Delta_{k_j}{}^{dN(k_j)}\prod_{\ell \neq k_j}\exp\{-\lambda(t_\ell \mid H_\ell)\Delta_\ell\} + o(\Delta_*)$$

$$= \exp\left\{\sum_{j=1}^{J} \log \lambda(t_{k_j} \mid H_{k_j}) dN(t_{k_j}) - \sum_{\ell \neq k_j} \lambda(t_\ell \mid H_\ell)\Delta_\ell \right\}$$

$$\times \exp\left\{\sum_{j=1}^{J} \log \Delta_{k_j} \right\} + o(\Delta_*), \tag{4.7}$$

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

$$p(N_{0,T}) = \lim_{\Delta_* \to 0} \frac{\exp\left\{\sum_{j=1}^{J} \log \lambda(t_{k_j} \mid H_{k_j}) dN(t_{k_j}) - \sum_{\ell \neq k_j} \lambda(t_\ell \mid H_\ell)\Delta_\ell \right\} \exp\left\{\sum_{j=1}^{J} \log \Delta_{k_j} \right\} + o(\Delta_*)}{\prod_{j=1}^{J} \Delta_j}$$

$$= \exp\left\{\int_0^T \log \lambda(u \mid H_u) dN(u) - \int_0^T \lambda(u \mid H_u) du \right\}. \tag{4.8}$$

Q.E.D.

Theorem 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.[1,5,6,7,8,10] That is, when formulated in terms of the conditional intensity function, all point process likelihoods have the form given in Eq. 4.8. The approximate probability density expressed in terms of the conditional intensity function (Eq. 4.7e) was given in [11].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.

4.2. An alternative derivation of the joint probability density of a spike train

This derivation is based on [1,12]. The sample path probability density is

$$p(u_1, u_2, ..., u_n \cap u_{n+1} > T) = \prod_{k=1}^{n} p(u_k \mid u_1, ..., u_{k-1}) \times \Pr(\textit{no spikes in } (u_n, T] \mid u_1, ..., u_n)$$

$$= \prod_{k=1}^{n} p(t_k \mid H_k)[1 - P(T \mid H_{n+1})], \tag{4.9}$$

where $[1 - P(T \mid H_{n+1})] = \int_{t_n}^{T} p(u \mid H_{n+1}) du$. From Eq. 3.2 we have for $t > t_{k-1}$,

$$r(t \mid H_k) = \frac{p(t \mid H_k)}{1 - P(t \mid H_k)}, \tag{4.10}$$

and

$$\lambda(t \mid H_k) = \begin{cases} r(t) & 0 < t < t_1 \\ r(t \mid H_k) & t_{k-1} < t < t_k \end{cases}. \tag{4.11}$$

Integrating both sides of Eq. 4.10 from t_{k-1} to t gives

$$-\log[1 - P(t \mid H_k)] = \int_{t_{k-1}}^{t} \lambda(u)du, \tag{4.12}$$

and after exponentiating both sides we obtain

$$[1 - P(t \mid H_k)] = \exp\left\{-\int_{t_{k-1}}^{t} \lambda(u \mid H_k)du\right\}. \tag{4.13}$$

Rearranging Eq. 4.10 and using Eq. 4.13 we have

$$\begin{aligned} p(t_k \mid H_k) &= r(t \mid H_k)[1 - P(t \mid H_k)] \\ &= \lambda(t \mid H_k)\exp\left\{-\int_{t_{k-1}}^{t} \lambda(u \mid H_u)du\right\}, \end{aligned} \tag{4.14}$$

which is the first part of the desired result. Returning to Eq. 4.9 and using Eq. 4.14, we find that the joint probability density of the spike train is

$$p(t_1, t_2, \ldots, t_n \cap t_{n+1} > T) = \prod_{k=1}^{n} p(t_k \mid H_k)[1 - P(T \mid H_{n+1})]$$

$$= \prod_{k=1}^{n} \lambda(t_k \mid H_k)\exp\left\{-\int_{t_{k-1}}^{t_k} \lambda(u \mid H_u)du\right\}[1 - P(T \mid H_{n+1})]$$

$$= \prod_{k=1}^{n} \lambda(t_k)\exp\left\{-\int_{0}^{T} \lambda(u)du\right\}$$

$$= \exp\left\{\int \log\lambda(u \mid H_u)dN(u) - \int_{0}^{T} \lambda(u \mid H_u)du\right\}. \tag{4.15}$$

5. Special point process models

Thus far, we have described how elementary stochastic dynamical systems models lead to interspike interval models (Section 2), developed a characterization of point processes in terms of the conditional intensity function (Section 3) and derived the joint probability density of a

point process in terms of the conditional intensity function (Section 4). In this section we describe some special point process models. These are the Poisson process, renewal processes and stationary point process models. This material follows [13 14, 15].

A counting process is said to have *independent increments* if the number of events which occur in disjoint time intervals is independent. A counting process is said to have *stationary increments* if the distribution of the number of events which occur in any time interval depends only on the length of the interval. That is, the process has stationary increments if the number of events in the interval (t_1+s, t_2+s) (i.e. $N(t_2+s)-N(t_1+s)$) has the same distribution as the number of events in the interval (t_1, t_2) (i.e. $N(t_2)-N(t_1)$) for all $t_1 < t_2$ and $s > 0$. Independent increments and stationary increments are very strong conditions that are rarely satisfied exactly for neural processes since history dependence of the spiking activity is the rule rather than the exception.

5.1. Poisson processes

5.1.1. Axioms for a Poisson process

A counting process $N(t) \quad t \geq 0$ is a Poisson process with rate $\lambda, \quad \lambda > 0$, if

i) $N(0) = 0$.

ii) The process has independent increments.

iii) $\Pr(N(t+\Delta) - N(t) = 1) = \lambda\Delta + o(\Delta)$.

iv) $\Pr(N(t+\Delta) - N(t) \geq 2) = o(\Delta)$.

The probability mass function is

$$p((N(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \tag{5.1}$$

for $k = 0, 1, 2, 3, \ldots$. $E[N(t)] = \lambda t$ and $Var[N(t)] = \lambda t$. There are many approaches to deriving this probability mass function. The probability mass function can be derived from assumptions i)—iv) by using the Poisson approximation to the binomial (See Section 8).

5.1.2. Interevent and waiting time probability densities for a Poisson process

To construct the interspike interval probability density for the Poisson process, we note that

$$\Pr(T > t) = \Pr(N(t) = 0) = e^{-\lambda t}, \tag{5.2}$$

or

$$\Pr(t \leq T) = 1 - e^{-\lambda t}. \tag{5.3}$$

If we differentiate Eq. 5.3, we find that the probability density is the exponential probability density

$$p(t) = \lambda e^{-\lambda t}, \tag{5.4}$$

for $t > 0$ and $\lambda > 0$. Given the assumption that the interspike interval probability density of a counting process is exponential, it is possible to show that this implies that the counting process is a Poisson process.

The waiting time probability density is the probability density of the time until the occurrence of the kth event. To construct the waiting time probability density, we consider $s_k = \sum_{i=1}^{k} t_i$. Using the properties of a Poisson process we obtain

$$\Pr(t < s_k < t+\Delta) = \Pr(N(t) = k-1 \cap \text{ an event in } (t, t+\Delta)) + o(\Delta)$$

$$= \Pr(N(t) = k-1)\Pr(\textit{one event in } (t, t+\Delta) \mid N(t) = 1) + o(\Delta)$$

$$\Pr(t < s_k < t+\Delta) \quad = \Pr(N(t) = k-1)\lambda\Delta + o(\Delta)$$

$$= \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} \lambda\Delta . \tag{5.5}$$

Hence,

$$p_{s_k}(t) = \lim_{\Delta \to 0} \frac{\Pr(t < s_k < t+\Delta)}{\Delta} = \frac{\lambda^k t^{k-1} e^{-\lambda t}}{(k-1)!} . \tag{5.6}$$

The waiting time probability density is, as we saw in Section 2, the gamma distribution with parameters n and λ. The conditional intensity function is

$$\lambda(t) = \frac{p(t)}{1 - \int_0^t p(u)du}$$

$$= \frac{\lambda e^{-\lambda t}}{1 - \int_0^t \lambda e^{-\lambda u} du}$$

$$= \lambda. \tag{5.7}$$

The joint probability density of the spike train (Eq. 4.15)

$$p(u_1, \ldots\ldots, u_n)\Pr(u_{n+1} > T - u_n) = \prod_{i=1}^{n} p(u_i \mid \lambda)\exp(-\lambda(T - u_n))$$

$$= \lambda^n \exp(-\lambda(\sum_{i=1}^{n}(u_i - u_{i-1}) - (T - u_n))$$

$$= \lambda^n \exp(-\lambda T). \tag{5.8}$$

There are three ways to generalize a simple Poisson process in order to construct more complex point process models. These are: 1) use a Poisson process with non-stationary increments, which leads to an inhomogeneous Poisson process; 2) chose an interevent probability model other than the exponential probability density which leads to a renewal process; and 3) allow the conditional intensity function to have history dependence which leads to the general class of point process models discussed in Sections 3 and 4. We examine these first two extensions next.

5.1.3. Inhomogeneous Poisson process

The following axioms define an inhomogeneous Poisson process.

i) $N(0) = 0$.

ii) $\{N(t),\ t \geq 0\}$ is a counting process with independent increments.

iii) $\Pr\{N(t+\Delta) - N(t) \geq 2\} = o(\Delta)$.

iv) $\Pr\{N(t+\Delta) - N(t) = 1\} = \lambda(t)\Delta + o(\Delta)$.

Given these axioms, it is straight forward to show that the probability mass function of the inhomogeneous Poisson process is

$$\Pr(N(t) = k) = \frac{[\Lambda(t)]^k e^{-\Lambda(t)}}{K!}, \tag{5.9}$$

where $\Lambda(t) = \int_0^t \lambda(u)du$. The interevent time probability density of the inhomogeneous Poisson can be derived by noting that

$$\Pr(T > t) = \Pr(N(t) = 0) = \exp\left\{-\int_0^t \lambda(u)du\right\} \tag{5.10}$$

or

$$\Pr(T \leq t) = 1 - \exp\left\{-\int_0^t \lambda(u)du\right\}. \tag{5.11}$$

and hence, by differentiating with respect to t we obtain

$$p(t) = \lambda(t)\exp\left\{-\int_0^t \lambda(u)du\right\}. \tag{5.12}$$

For an alternative derivation of this probability density we take $p(z) = e^{-z}$ and let $z = \int_0^t \lambda(u)du$. We have $p(t)dt = f(z)dz$ and hence $p(t) = p(z)\frac{dz}{dt}$ which is the standard change-of-variables formula. Now $\frac{dz}{dt} = \lambda(t)$ and hence $p(t) = \lambda(t)e^{-z} = \lambda(t)\exp\left\{-\int_0^t \lambda(u)du\right\}$.We have inverted the time-rescaling theorem (see Section 6).

From the arguments used to construct the waiting time probability density for the simple Poisson process, it is easy to show that the waiting time probability density for the inhomogeneous Poisson process is

$$p_{s_k}(t) = \frac{\lambda(t)\Lambda(t)^{k-1}\exp(-\Lambda(t))}{\Gamma(k)} \quad . \tag{5.13}$$

From **Theorem 1**, it follows that the joint probability density of the spike times is

$$\begin{aligned} p(u_1,\ldots\ldots,u_n \cap u_{n+1} > T) &= p(u_1,\ldots,u_n)\Pr(u_{n+1} > T \mid u_1,\ldots,u_n) \\ &= \prod_{i=1}^{n}\lambda(u_i)\exp\left\{-\int_{u_{i-1}}^{u_i}\lambda(u)du\right\}\exp\left\{-\int_{u_n}^{T}\lambda(u)du\right\} \\ &\quad \prod_{i=1}^{n}\lambda(u_i)\exp\left\{-\int_0^T\lambda(u)du\right\}, \end{aligned} \tag{5.14}$$

because

$$\begin{aligned} \Pr(u_{n+1} > T \mid u_1,\ldots,u_n) &= \Pr(u_{n+1} > T) = \Pr(\text{No spike in } (u_n, T] \\ &= \exp\left\{-\int_{s_n}^{t}\lambda(u)du\right\}. \end{aligned} \tag{5.15}$$

To show that the rate function and the conditional intensity function are the same for an inhomogeneous Poisson model we substitute Eq. 5.12 in Eq. 3.2 to obtain

$$\begin{aligned} \lambda(t \mid H_k) &= \frac{\lambda(t)\exp\left\{-\int_{t_{k-1}}^{t}\lambda(u)du\right\}}{1-P(t \mid H_k)} \\ &= \frac{\lambda(t)\exp\left\{-\int_{t_{k-1}}^{t}\lambda(u)du\right\}}{\exp\left\{-\int_{t_{k-1}}^{t}\lambda(u)du\right\}} \\ &= \lambda(t). \end{aligned} \tag{5.16}$$

Therefore the conditional intensity function for the inhomogeneous Poisson process is the rate function.

Given an observation interval $(0,T]$ and a Poisson process with intensity parameter $\lambda(t)$, we consider the distribution of the Poisson events conditional on $N(T) = n$. That is

$$\Pr(u_1,u_2,\ldots,u_n \mid N(T)=n) = \frac{\Pr(u_1,\ldots,u_n \cap N(T)=n)}{\Pr(N(T)=n)}$$

$$= \frac{\prod_{k=1}^{n}\lambda(u_k)\exp\left\{-\int_0^T \lambda(u)du\right\}}{\frac{\left[\int_0^T \lambda(u)du\right]^n}{n!}}\exp\left\{-\int_0^T \lambda(u)du\right\}$$

$$= \frac{n!\prod_{k=1}^{n}\lambda(u_k)}{\left[\int_0^T \lambda(u)du\right]^n}. \tag{5.17}$$

This is the joint distribution of the ordered observation from the probability density

$$p(t) = \frac{\lambda(t)}{\int_0^T \lambda(u)du}, \tag{5.18}$$

where $t\varepsilon(0,T]$. This result is very relevant for neuroscience data analysis because a common practice is to use the normalized peristimulus time histogram (PSTH) as a probability density. The above result gives conditions under which this is reasonable.

5.2. Renewal processes

A renewal process is a point process in which the interevent interval are independent and drawn from the same probability density. More specifically, let T_i be independent, identically distributed interevent times from the probability density $p(t)$. Let $s_r = \sum_{i=1}^{r} T_i$. The process s_r is a *renewal process*. The gamma and inverse Gaussian probability densities derived in Section 2 are renewal processes. Since we have defined the interevent (interspike) interval distribution we go back and find the counting process probability mass function.

5.2.1. Counting process associated with a renewal process

The counting process $N(t)$, is the number of events in $(0,t]$. Now $N(t)<r$ if and only if $s_r > t$ or, $\Pr(N(t)<r) = \Pr(s_r > t) = 1-P_r(t)$. Therefore,

$$\Pr(N(t)\geq r) = \Pr(N(t)=r \bigcup N(t)>r)$$

$$= \Pr(N(t)=r \bigcup N(t)\geq r+1)$$

$$= \Pr(N(t)=r) + \Pr(N(t)\geq r+1).$$

Therefore, we have

$$\Pr(N(t)=r) = \Pr(N(t)\geq r) - \Pr(N(t)\geq r+1)$$

$$= P_r(t) - P_{r+1}(t). \tag{5.19}$$

where $P_r(t)$ is the probability distribution function associated with the r-fold convolution of $p(t)$. While Eq. 5.19 gives the probability mass function for the counting process associated with any renewal process, it can only be evaluated explicitly in specific cases.

5.2.2. Asymptotic distribution of $N(t)$ for large t.

Because Eq. 5.19 is a challenge to evaluate, asymptotic approximations are frequently used. It can be shown [14,15] that for large t the distribution of $N(t)$ may be approximated as a Gaussian random variable with mean and variance defined by

$$E[N(t)] = t\mu^{-1} \quad Var[N(t)] = \sigma^2\mu^{-3}t \tag{5.20}$$

where

$$\mu = E[t_i] \quad \sigma^2 = Var[t_i]. \tag{5.21}$$

5.3. Stationary process

We mention briefly the class of stationary point process models [14,15] and show how its properties can be analyzed in terms of the conditional intensity function. Under the assumption that the point process is stationary, we have that

$$\begin{aligned} E[dN(t)] &= \lambda(t \mid H_t) \\ &= \lambda dt \end{aligned} \tag{5.22}$$

and that the variance is given by

$$Var[dN(t)] = \sigma^2 dt. \tag{5.23}$$

5.3.1. Covariance function. To find the covariance function for a stationary point process we consider

$$\operatorname{cov}(dN(t+v), dN(t)) = \Pr(dN(t)=1 \cap dN(t+v)=1)$$

$$= \Pr(dN(t)=1)\Pr(dN(t+v)=1 \mid dN(t)=1) \tag{5.24}$$

$$= \lambda\Delta t\lambda(v)\Delta t = \lambda\lambda(v)(\Delta t)^2,$$

where

$$\lambda(v) = \lim_{\Delta\to 0} \frac{\Pr(\text{event } (t+v, t+v+\Delta] \mid \text{ event at } t)}{\Delta} \tag{5.25}$$

is the conditional intensity function of a stationary point process model. Hence, we have

$$Cov(dN(t+v), dN(t)) = \gamma(v)(\Delta t)^2, \tag{5.26}$$

where $\gamma(v) = \lambda(\lambda(v) - \lambda)$ is the covariance function of the process or more generally

$$\gamma(v) = \lambda(\lambda(v) - \lambda) + \sigma^2 \delta(v), \tag{5.27}$$

where $\delta(v)$ is the Dirac function and allows us to consider the variance of the process as part of the same expression.

5.3.2. Spectral density function

Taking the Fourier transform of the covariance function we find that the spectral density function is

$$\phi(\omega) = \frac{\lambda}{2\pi}\int_{-\infty}^{\infty}(\lambda(v) - \lambda)e^{-i\omega v}dv + \frac{\sigma^2\delta(v)}{2\pi}. \tag{5.28}$$

Unlike with stationary Gaussian processes, it is possible to have two different stationary point process models with the same spectral density.

6. The time-rescaling theorem

This result, originally due to [16] and [17], states that every point process with a conditional intensity function maps into a Poisson process with unit rate. In addition to being an interesting theoretical result, it has important implications for assessing goodness-of-fit for point process models of neural spike trains.

We begin by restating the result established in the previous section. The joint probability density of exactly n event times in $(0,T]$ is

$$\begin{aligned} p(u_1, u_2, \ldots, u_n \cap N(T) = n) &= p(u_1, u_2, \ldots, u_n \cap t_{n+1} > T) \\ &= p(u_1, u_2, \ldots, u_n)\Pr(u_{n+1} > T \mid u_1, u_2, \ldots, u_n) \qquad (6.1) \\ &= \prod_{k=1}^{n}\lambda(u_k \mid H_{t_k})\exp\left\{-\int_{u_{k-1}}^{u_k}\lambda(u \mid H_u)du\right\}\exp\left\{-\int_{u_n}^{T}\lambda(u \mid H_u)du\right\}, \end{aligned}$$

where

$$p(u_1, u_2, \ldots, u_n) = \prod_{k=1}^{n}\lambda(u_k \mid H_{t_k})\exp\left\{-\int_{u_{k-1}}^{u_k}\lambda(u \mid H_u)du\right\} \tag{6.2}$$

$$\Pr(t_{n+1} > T \mid t_1, t_2, \ldots, t_n) = \exp\left\{-\int_{t_n}^{T}\lambda(u \mid H_u)du\right\}. \tag{6.3}$$

and $t_0 = 0$. We can now state and prove the time-rescaling theorem.

6.1. An elementary proof of the time-rescaling theorem.

This proof follows closely [18].

Theorem 2. Let $0 < u_1 < u_2 <, \ldots, < u_n < T$ be a realization from a point process with a conditional intensity function $\lambda(t \mid H_t)$ satisfying $0 < \lambda(t \mid H_t)$ for all $t \in (0,T]$. Define the transformation

$$\Lambda(u_k) = \int_0^{u_k} \lambda(u \mid H_u) du \tag{6.4}$$

for $k = 1, \ldots, n$, and assume $\Lambda(t) < \infty$ with probability one for all $t \in (0,T]$. Then the $\Lambda(u_k)$'s are a Poisson process with unit rate.

Proof: Let $\tau_k = \Lambda(u_k) - \Lambda(u_{k-1})$ for $k = 1, \ldots, n$ and set $\tau_T = \int_{u_n}^{T} \lambda(u \mid H_u) du$. To establish the result it suffices to show that the τ_ks are independent, identically distributed exponential random variables with mean one. Because the τ_k transformation is one-to-one and $\tau_{n+1} > \tau_T$ if and only if $u_{n+1} > T$, the joint density of the τ_k's is

$$p(\tau_1, \tau_2, \ldots, \tau_n \cap \tau_{n+1} > \tau_T) = p(\tau_1, \ldots, \tau_n) \Pr(\tau_{n+1} > \tau_T \mid \tau_1, \ldots, \tau_n). \tag{6.5}$$

We evaluate each of the two terms on the right side of Eq. 6.5. The following two events are equivalent

$$\{\tau_{n+1} > \tau_T \mid \tau_1, \ldots, \tau_n\} = \{u_{n+1} > T \mid u_1, u_2, \ldots, u_n\}. \tag{6.6}$$

Hence

$$\begin{aligned} \Pr(\tau_{n+1} > \tau_T \mid \tau_1, \tau_2, \ldots, \tau_n) &= \Pr(u_{n+1} > T \mid u_1, u_2, \ldots, u_n) \\ &= \exp\left\{-\int_{u_n}^{T} \lambda(u \mid H_{u_n}) du\right\} \\ &= \exp\{-\tau_T\}, \end{aligned} \tag{6.7}$$

where the last equality follows from the definition of τ_T. By the multivariate change-of-variable formula [19]

$$p(\tau_1, \tau_2, \ldots, \tau_n) = \mid J \mid p(u_1, u_2, \ldots, u_n \cap N(u_n) = n), \tag{6.8}$$

where J is the Jacobian of the transformation between $u_j,\ j = 1, \ldots, n$ and $\tau_k, k = 1, \ldots, n$. Because τ_k is a function of $u_1, \ldots, u_k$, J is a lower triangular matrix and its determinant is the product of its diagonal elements defined as $\mid J \mid = \mid \prod_{k=1}^{n} J_{kk} \mid$. By assumption $0 < \lambda(t \mid H_t)$ and, by Eq. 6.4 and the definition of τ_k the mapping of u into τ is one-to-one. Therefore, by the inverse differentiation theorem [20] the diagonal elements of J are

$$J_{kk} = \frac{\partial u_k}{\partial \tau_k} = \lambda(u_k \mid H_{u_k})^{-1}. \tag{6.9}$$

Substituting $|J|$ and Eq. 6.2 into Eq. 6.8 yields

$$p(\tau_1,\tau_2,\ldots,\tau_n)=\prod_{k=1}^{n}\lambda(u_k\mid H_{u_k})^{-1}\prod_{k=1}^{n}\lambda(u_k\mid H_{u_k})\exp\left\{-\int_{u_{k-1}}^{u_k}\lambda(u\mid H_u)du\right\}$$

$$=\prod_{k=1}^{n}\exp\{-[\Lambda(u_k)-\Lambda(u_{k-1})]\} \tag{6.10}$$

$$=\prod_{k=1}^{n}\exp\{-\tau_k\}.$$

Substituting Eq. 6.10 and Eq. 6.7 into Eq. 6.5 yields

$$p(\tau_1,\tau_2,\ldots,\tau_n\cap\tau_{n+1}>\tau_T)=p(\tau_1,\ldots,\tau_n)\Pr(\tau_{n+1}>\tau_T\mid\tau_1,\ldots,\tau_n)$$

$$=\left(\prod_{k=1}^{n}\exp\{-\tau_k\}\right)\exp\{-\tau_T\}, \tag{6.11}$$

which establishes the result.

The time-rescaling theorem generates a history-dependent rescaling of the time axis that converts a point process into a Poisson process with a unit rate (**Fig. 5**).

Figure 5 about here.

6.2. The time-rescaling theorem: assessing model goodness-of-fit

We may use the time-rescaling theorem to construct goodness-of-fit tests for a 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

$$\tau_k=\Lambda(u_k)-\Lambda(u_{k-1}). \tag{6.12}$$

If the model is correct then, according to the theorem, the $\tau_k s$ are independent exponential random variables with mean 1. If we make the further transformation

$$z_k=1-e^{-\tau_k}, \tag{6.13}$$

then $z_k s$ are independent uniform random variables on the interval [0,1). Because the transformations in Eqs. 6.12 and 6.13 are both one-to-one, any statistical assessment that measures agreement between the $z_k s$ and a uniform distribution directly evaluates how well the original model agrees with the spike train data [21]. Here we present two methods: Kolmogorov-Smirnov tests and quantile-quantile plots.

6.2.1. Kolmogorov-Smirnov test. To construct the Kolmogorov-Smirnov test we first order the $z_k s$ from smallest to largest, denoting the ordered values as $z_{(k)} s$. We then plot the values of the

cumulative distribution function of the uniform density defined as $b_k = \frac{k-\frac{1}{2}}{n}$ for $k = 1, \ldots, n$ against the $z_{(k)}s$. If the model is correct, then the points should lie on a 45° line [22]. Confidence bounds for the degree of agreement between the models and the data may be constructed using the distribution of the Kolmogorov-Smirnov statistic [22]. For moderate to large sample sizes the 95% (99%) confidence bounds are well approximated as $b_k \pm 1.36/n^{1/2}$ ($b_k \pm 1.63/n^{1/2}$) [22]. We term such a plot a Kolmogorov-Smirnov (KS) plot.

6.2.2. Quantile-quantile plot. Another approach to measuring agreement between the uniform probability density and the $z_k s$ is to construct a quantile-quantile (Q-Q) plot [22]. In this display we plot the quantiles of the uniform distribution, denoted here also as the $b_k s$, against the $z_{(k)}s$. As in the case of the KS plots, exact agreement occurs between the point process model and the experimental data if the points lie on a 45° line. Pointwise confidence bands can be constructed to measure the magnitude of the departure the plot from the 45° line relative to chance. To construct pointwise bands we note that if the $\tau_k s$ are independent exponential random variables with mean 1 and the $z_k s$ are thus uniform on the interval $(0,1]$ then, each $z_{(k)}$ has a beta probability density with parameters k and $n-k+1$ defined as

$$p(z \mid k, n-k+1) = \frac{n!}{(n-k)!(k-1)!} z^{k-1}(1-z)^{n-k}. \qquad (6.14)$$

for $0 < z \le 1$ [21]. We set the 95% confidence bounds by finding the 2.5th and 97.5th quantiles of the cumulative distribution associated with Eq. 2.18 for $k = 1, \ldots, n.$ These exact quantiles are readily available in many statistical software packages. For moderate to large spike train data series, a reasonable approximation to the 95% (99%) confidence bounds is given by the Gaussian approximation to the binomial probability distribution as $z_{(k)} \pm 1.96[z_{(k)}(1-z_{(k)})/n]^{1/2}$ ($z_{(k)} \pm 2.575[z_{(k)}(1-z_{(k)})/n]^{1/2}$). These local confidence bounds for the Q-Q plots based on the beta distribution and the Gaussian approximation are, to our knowledge, new.

In general, the KS confidence intervals will be wider than the corresponding Q-Q plot intervals. To see this, it suffices to compare the widths of the two intervals using their approximate formulae for large n. From the Gaussian approximation to the binomial, the maximum width of the 95% confidence interval for the Q-Q plots occurs at the median, i.e., $z_{(k)} = 0.50$, and is $2[1.96/(4n)^{1/2}] = 1.96n^{-1/2}$. For n large, the width of the 95% confidence intervals for the KS plots is $2.72n^{-1/2}$ at all quantiles. The KS confidence bounds consider the maximum discrepancy from the 45° line along all quantiles; the bands show the discrepancy that would be exceeded 5% of the time by chance if the plotted data were truly uniformly distributed. The Q-Q plot confidence bounds consider the maximum discrepancy from the 45° line for each quantile separately. These pointwise 95% confidence bounds mark the amount by which each value $z_{(k)}$ would deviate from the true quantile 5% of the time purely by chance. The KS bounds are broad because they are based on the joint distribution of all n deviations and they consider the distribution of the largest of these deviations. The Q-Q plot bounds are narrower because they measure the deviation at each quantile separately. Used together, the two plots help approximate

upper and lower limits on the discrepancy between a proposed model and a spike train data series.

6.2.3. Normalized point process residuals. Let $0 < t_1^* < \cdots < t_K^* < T$ be a coarse partition of the observation interval $(0T]$. Let N_k^* denote the number of spikes in the interval $(t_{k-1}^*, t_k^*]$. Define the point process residuals

$$r_k = \frac{N_k^* - \Lambda(\Delta_k^*)}{[\Lambda(\Delta_k^*)]^{\frac{1}{2}}},$$

for $k = 1, \ldots, K$. The point process residuals have zero expectation and variance 1. A plot of the point process residuals against time can reveal systematic temporal structures in model lack of fit.

As a second application of the time-rescaling theorem we discuss in Section 7 an approach to simulating point processes based on the time-rescaling theorem.

7. Simulation of point processes

Lewis and Shedler [23] developed an efficient algorithm for simulating a Poisson process termed thinning. Ogata [24] developed extension to the general point processes using the conditional intensity function. The algorithms are as follows.

7.1. Thinning algorithm 1. Given a regular point process $N(t)$ on $(0,T]$ with a bounded intensity function $\lambda(t)$. That is, for all $t\varepsilon(0,T]$ there exists λ such that $\lambda(t) \le \lambda$. To simulate a random sample from $N(t)$ use the following two stage algorithm:

A. Draw a Poisson Sample

1. Draw u_i from $u(0,1)$
2. Compute $w_i = -\frac{\log(1-u_i)}{\lambda}$
3. Compute $t_i = t_{i-1} + w_1$
4. If $t_i \ge T$ stop else $i = i+1$ go to 1 (7.1)
5. $N(T) = i$

B. Thinning Algorithm

0. Set $i = 1$ and $j = 1$
1. Draw v_i from $U(0,1)$
2. Compute $\lambda(t_i)/\lambda$
3. If $\lambda(t_i)/\lambda \le v_i$ accept $t_j = t_i$; $j = j+1$
4. If $i = N(T)$ stop else $i = i+1$; go to 1.

The t_j's are a random sample from $N(T)$.

The proof is straightforward. It suffices to note that

$$\Pr(\text{accepting } t_i \text{ in } [t,t+\Delta t)) = \Pr(t_i \text{ in } [t,t+\Delta t))\Pr(\text{accepting } t_i \mid t_i \varepsilon [t,t+\Delta t))$$

$$= \lambda \Delta t \frac{\lambda(t_i)}{\lambda} \tag{7.2}$$

$$= \lambda(t_i)\Delta t.$$

Hence t_i is a sample from the point process with conditional intensity function $\lambda(t)$.

Figure 6 about here.

7.2. Simulating a multivariate point process model.
Although we have not formally defined a multivariate point process, it is roughly speaking a set of univariate point processes whose probability structure is characterized by a vector valued joint intensity function. Such models will be necessary to accurately describe multiple single unit neural spike train data. See [25] for a discussion of these models.

Theorem 3. [24] Consider an m-dimensional multivariate point process model on an interval $(0,T]$ with joint intensity function $\lambda_t = (\lambda_t^1,...,\lambda_t^m)$. Suppose that there is a homogeneous (univariate) Poisson process with rate λ^* such that for all t

$$\sum_{k=1}^{m} \lambda_t^k \le \lambda^* \tag{7.3}$$

and

$$\lambda_t^0 = \lambda^* - \sum_{k=1}^{m} \lambda_t^k \ . \tag{7.4}$$

Let $0 < t_{1^*}, < t_{2^*}, ..., < t_{n^*} \le T$ be the points of the Poisson process with intensity λ^*. For each of the points accept t_{j^*} to be a sample point from the k^{th} component of λ^* with probability λ_t^k / λ^*. Then the accepted points are a sample from a multivariate point process with parameter λ_t.

Proof: The proof is again straightforward. It suffices to note that

$$\Pr(\text{accepting } t_{j^*} \text{ in } [t,t+\Delta t)) = \Pr(t_{j^*} \text{ from } \lambda_t^k \text{ in } [t,t+\Delta t))\Pr(\text{accepting } t_{j^*} \mid t_{j^*} \in [t,t+\Delta t))$$

$$= \lambda^* \Delta t \frac{\lambda_t^k(t_{j^*})}{\lambda^*}$$

$$= \lambda_t^k(t_{j^*})\Delta t. \tag{7.5}$$

Hence t_{j^*} is a sample from the point process with conditional intensity function $\lambda_t^k(t_{j^*})$.

Notice that this probability is a multinomial probability, for each t_{j^*}. That is, there is probability that $\lambda_t^k(t_{j^*})$ that t_{j^*} will be assigned neuron k for $k = 1,....,m$ and probability λ_t^0 of being assigned to no neuron. Notice that as a consequence, that this procedure does not admit simultaneous spikes. In the proof in 24**Error! Bookmark not defined.**], the parent homogeneous Poisson process from which the candidate spike times are drawn is chosen to be piecewise continuous so that the bound can be chosen locally to make the algorithm more efficient.

7.3. Simulating a univariate point process by time-rescaling. As a second application of the time-rescaling theorem we describe how the theorem may be used to simulate a general point process.[18] The time-rescaling theorem provides a standard approach for simulating an inhomogeneous Poisson process from a simple Poisson process. The general form of the time-rescaling theorem suggests that any point process may be simulated from a Poisson process with unit rate by rescaling time with respect to the conditional intensity (rate) function of the process. Given an interval $(0,T]$ the simulation algorithm proceeds as follows:

1. Set $u_0 = 0$; Set $k = 1$.
2. Draw τ_k an exponential random variable with mean 1.
3. Find u_k as the solution to $\tau_k = \int_{u_{k-1}}^{u_k} \lambda(u \mid u_0, u_1, ..., u_{k-1})du$. (7.6)
4. If $u_k > T$ then stop.
5. $k = k + 1$
6. Go to 2.

By using Eq. 7.6 a discrete version of the time-rescaling algorithm can be constructed as follows. Choose J large, and divide the interval $(0,T]$ into J bins each of width $\Delta = T/J$. For $k = 1,..., J$ draw a Bernoulli random variable u_k^* with probability $\lambda(k\Delta \mid u_1^*,...,u_{k-1}^*)\Delta$ and assign a spike to bin k if $u_k^* = 1$, and no spike if $u_k^* = 0$.

While in many instances there will be faster, more computationally efficient algorithms for simulating a point process, such as model based methods for specific renewal processes [26] and thinning algorithms, the algorithm in Eq. 7.6 is simple to implement. For an example of where this algorithm is crucial for point process simulation we consider the inhomogeneous inverse gamma model in [27]. Its conditional intensity function is infinite immediately following a spike if $\psi < 1$. If in addition, ψ is time-varying, i.e., $\psi = \psi(t) < 1$ for all t, then neither thinning nor standard algorithms for making draws from a gamma probability distribution may be used to simulate data from this model. Thinning fails because the conditional intensity function is not bounded and the standard algorithms for simulating a gamma model cannot be applied because

ψ is time-varying. In this case, the time-rescaling simulation algorithm may be applied as long as the conditional intensity function remains integrable as ψ varies temporally.

For algorithms to simulate specific renewal models see [26].

8. Poisson limit theorems

One method for deriving a Poisson process is as a limit of a binomial process when the number of attempts or trials, n, is large and the probability of success p is small. A second approach is a superposition of non-Poisson processes. We discuss both briefly.

8.1. Poisson approximation to the binomial probability mass function

Suppose x has a binomial distribution with parameters n and p, where n is large and p is small. Let $\lambda = np$

$$\Pr(x=k) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}$$

$$= \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \tag{8.1}$$

$$= \frac{n(n-1)\ldots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k}$$

Now for n large and p small

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \quad \frac{n(n-1)\ldots(n-k+1)}{n^k} \approx 1, \quad \left(1-\frac{\lambda}{n}\right) \approx 1 .$$

Hence for n large and p small we have

$$\Pr(x=k) \approx \frac{\lambda^k}{k!} e^{-\lambda} . \tag{8.2}$$

8.2 Superposition of point processes

Another way to construct a point process is to superpose a set of point processes. The question then arises as to what is the distribution of the superposed process. A case in neuroscience data analysis where this is relevant is in the construction of the peristimulus time histogram (PSTH). The PSTH is constructed by choosing a fixed reference point and superposing the spike trains across multiple run in an experiment with respect to this point. It is often assumed that the resultant distribution of spikes is Poisson. In general superposition of point processes will not lead to a composite process that is a Poisson process [14]. This assumption however, can be easily verified empirically by using the goodness-of-fit analysis based on the time-rescaling theorem discussed in Section 6**.** One case in which the resultant

process is Poisson is if the spiking activity is sparse as the number of component processes increases, in the spirit of the Poisson approximation to the binomial.

9. PROBLEMS

1. Let $w_1,...,w_n$ be independent, identically distributed gamma random variables where

$$f(w_i \mid \alpha_1\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} w_i^{\alpha-1} e^{-\beta w_i}$$

$w_i > 0, \alpha > 0, \beta > 0$ and $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$ Let $s_n = \sum_{i=1}^{n} w_i.$ Show that the probability density of

$$f(s_n \mid \alpha^*, \beta) = \frac{\beta^{\alpha^*}}{\Gamma(\alpha^*)} s_n^{\alpha^*-1} e^{-\beta s_n}$$

where $\alpha^* = n\alpha$

2. Show that if w has an inverse Gaussian probability density with parameters μ' and λ'. Verify that $E(w) = \mu'$ and $Var(w) = \mu'^3\lambda^{-1}$.

3. Let $w_1...,w_n$ be independent, identically distributed gamma random variables. Write down the joint probability density in terms of the interevent time probability density and in terms of the conditional intensity function. Show that while the interevent times are independent, the counting process increments are correlated.

4. If z has an exponential probability density with parameter λ, show that $u = 1 - e^{-z}$ is a uniform random variable on the interval $(0,1)$.

5. Consider a point process whose conditional intensity function is

$$\lambda(t \mid H_t) = \mu + \alpha \int_0^t e^{-\beta(t-u)} dN(u)$$

defined an interval $(0,T]$, where $\mu > 0,$ $\alpha > 0,$ and $\beta > 0.$ Write a thinning algorithm to simulate this point process model. How can you use the time-rescaling theorem and the $K-S$ plots to check the accuracy of your simulation? Write a continuous time and discrete time algorithm to simulate this model based on the time-rescaling theorem.

References

[1] D. Daley and Vere-Jones, D. *An Introduction to the Theory of Point Process (2nd ed.)* (Springer-Verlag, New York, 2003).

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

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

[4] G.L. Gerstein and B. Mandelbrot, J. Biophys **4** (1964) 41-68.

[5] E.N. Brown, R. Barbieri, U.T. Eden, L.M. Frank, in: *Computational Neuroscience: A Comprehensive Approach* (CRC, London, 2003).

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

[7] M. Jacobsen, *Statistical Analysis of Counting Processes* (Springer-Verlag, New York, 1982).

[8] P. Guttorp, *Stochastic Modeling of Scientific Data* (Chapman & Hall, London, 1995).

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

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

[11] D.R. Brillinger, Biol. Cyber. **59** (1988) 189-200.

[12] R. Barbieri, M.C. Quirk, L.M. Frank, M.A. Wilson and E.N. Brown, J. Neurosci. Meth. **105** (2001) 25-37.

[13] S. M. Ross, *Introduction to Probability Models* (6th ed.) (Academic, Press, San Diego, 1997).

[14] D.R. Cox and V. Isham, *Point Processes*, (Chapman and Hall, London, 1980).

[15] D.R. Cox and Miller, H.D. *The Theory of Stochastic Processes* (Methuen, London, 1965).

[16] P. Meyer *Démonstration Simplifiée d'un Théorème de Knight, Séminaire Probabilitié V. Lecture Notes in Mathematics* (Springer-Verlag, New York, 1969) 191-195.

[17] F. Panangelou, Trans. Amer. Math. Soc. **165** (1972) 483-506.

[18] E.N. Brown, R. Barbieri, V. Ventura, R.E. Kass, L.M. Frank, Neural Computation **14(2)** (2002) 325-46.

[19] S. Port, *Theoretical Probability for Applications* (Wiley, New York, 1994).

[20] M.H. Porter and C.B. Morrey, *A First Course in Real Analysis (2nd ed.)* (Springer- Verlag, New York, 1991).

[21] Y. Ogata, Journal of the American Statistical Association **83** (1988) 9-27.

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

[23] P.A.W. Lewis and G.S. Shedler, Naval Res. Logistics Quart. **26** (1978) 403-413.

[24] Y. Ogata, IEEE Transactions on Information Theory **27** (1981) 23-31.

[25] E.N. Brown, R.E. Kass, and P.M. Mitra, Nature Neuroscience (2004).

[26] B. Ripley, *Stochastic Simulation* (Wiley, New York, 1987).

[27]. R. Barbieri, M.C. Quirk, L.M. Frank, M.A. Wilson and E.N. Brown, J. Neurosci. Meth. **105** (2001) 25-37.

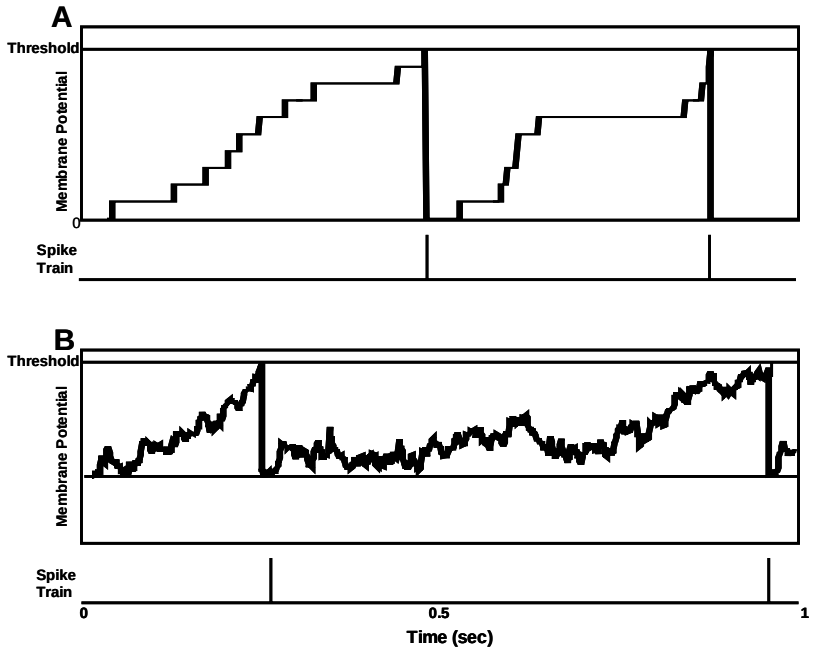

Figure 1. Voltage time course for A, a non-leaky integrate and fire neuron with inputs given as a Poisson point process with rate λ and input amplitude α. (Eq. 2.1) B, a non-leaky integrate-and-fire neuron with following Wiener process with drift (Eq. 2_). For both models when the membrane potential crosses threshold an action potential is discharged and the membrane voltage resets immediately to the resting potential.

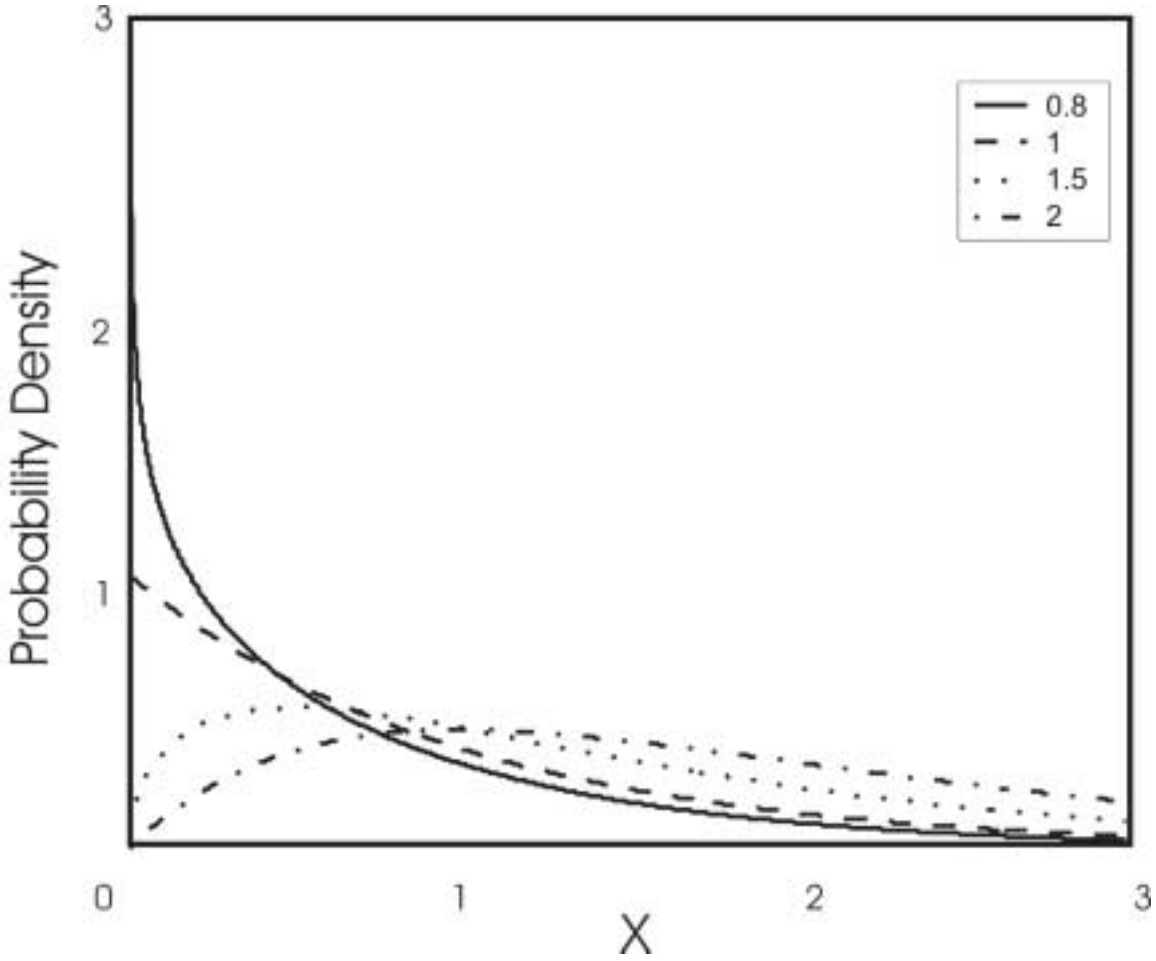

Figure 2. Examples of gamma probability densities for $\lambda = 1$, and $\alpha = 0.8$, $\alpha = 1.0$, $\alpha = 1.5$ and $\alpha = 2.0$.

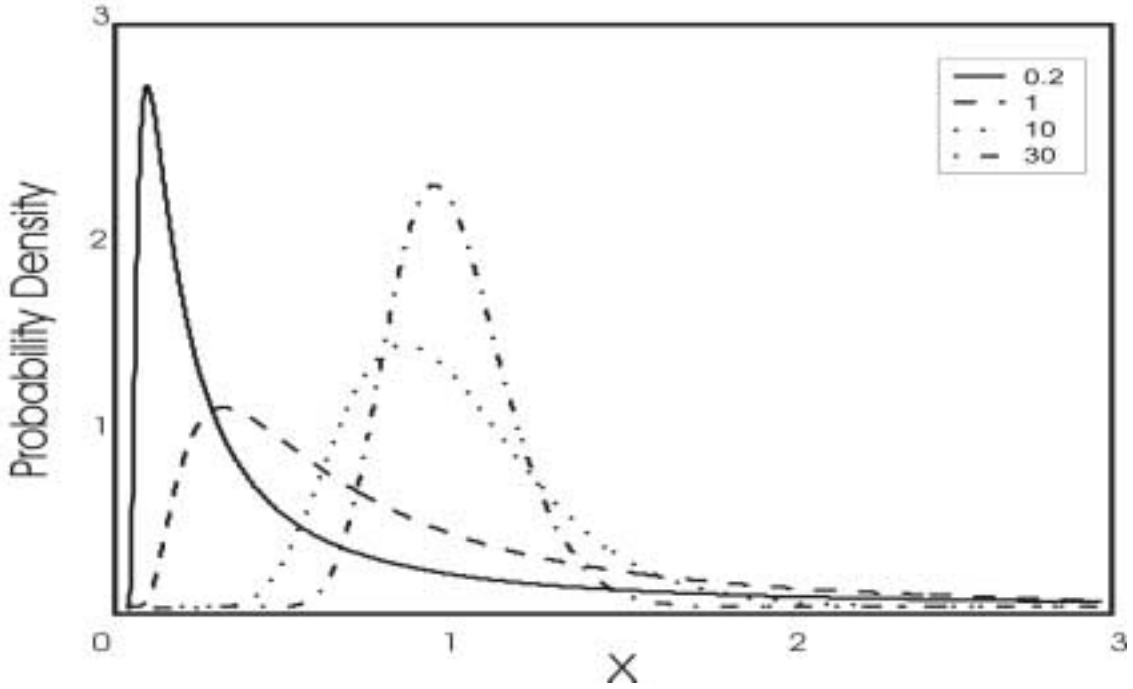

Figure 3. Examples of inverse Gaussian probability densities for $\lambda = 1$, and $\alpha = 0.8$, $\alpha = 1.0$, $\alpha = 1.5$ and $\alpha = 2.0$.

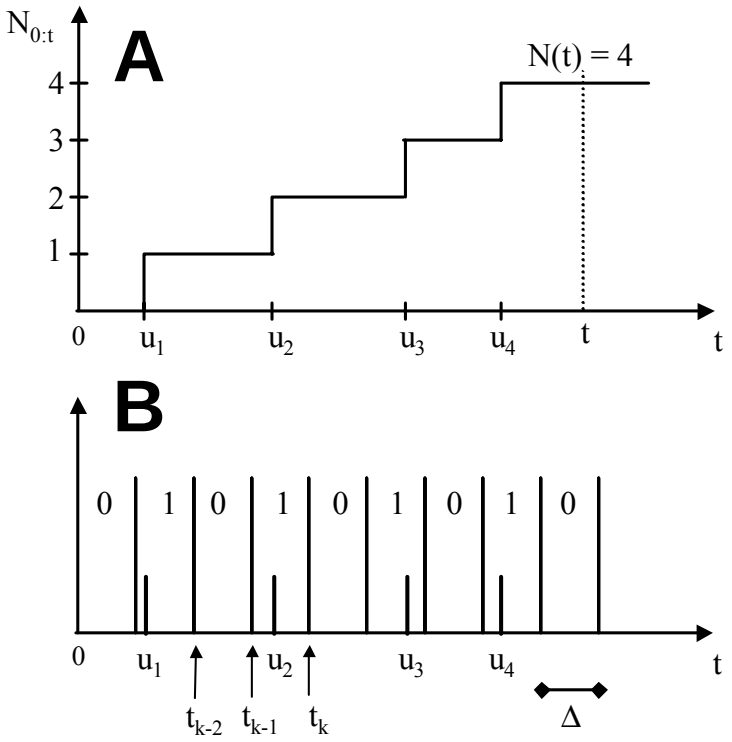

Figure 4. A. The construction of the sample path $N_{0:t}$ from the spike times u_1, u_2, u_3, u_4. At time t, $N_{0:t} = \{u_1, u_2, u_3, u_4 \cap N(t) = 4\}$ B. The discretization of the time axis to evaluate the probability of each spike occurrence or non-occurrence as a local Bernoulli process. By Eq. 3.3 the probability of the event u_2, i.e. a 1 between t_{k-1} and t_k, is $\lambda(t_k \mid H_k)\Delta$ whereas the probability of the event immediately prior to u_2, i.e. a 0 between t_{k-2} and t_{k-1}, is $1-\lambda(t_{k-1} \mid H_{k-1})\Delta$. In this plot we have taken $\Delta_k = \Delta$ for all $k = 1,...,K$ (reprinted and used with permission of CRC Press).

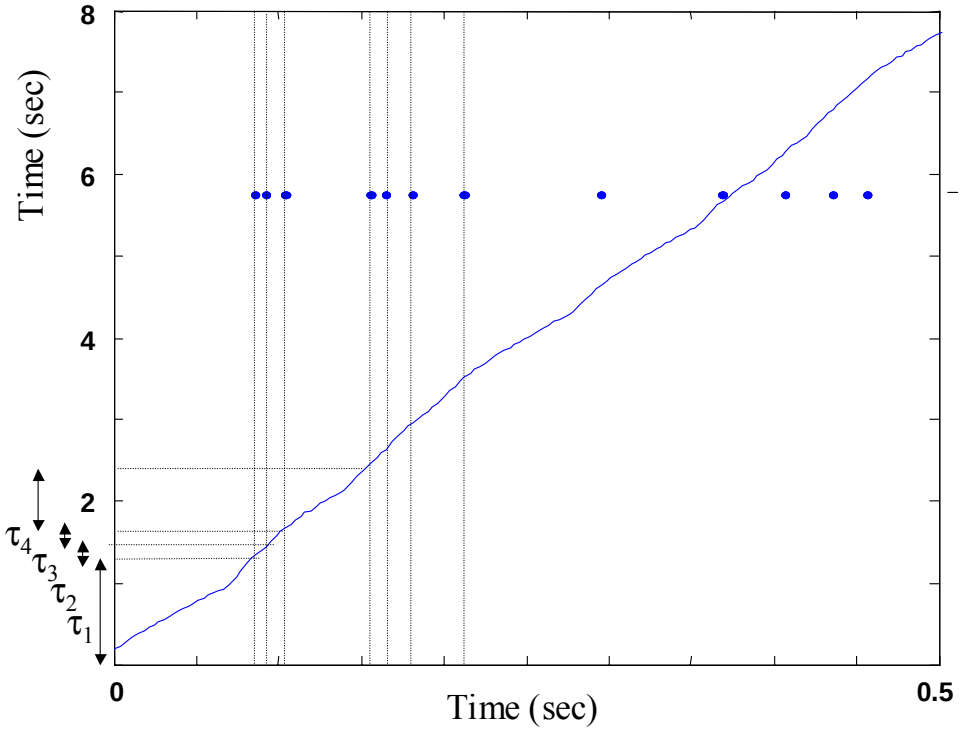

Figure 5. Illustration of the time rescaling theorem. The interspike intervals (x-axis) $t_k - t_{k-1}$ are transformed into the $\tau_k s$ (y-axis) using Eqs. 6.13. The integrated conditional intensity function is the solid line. By the time rescaling theorem the $\tau_k s$ are independent, identically distributed exponential random variables.

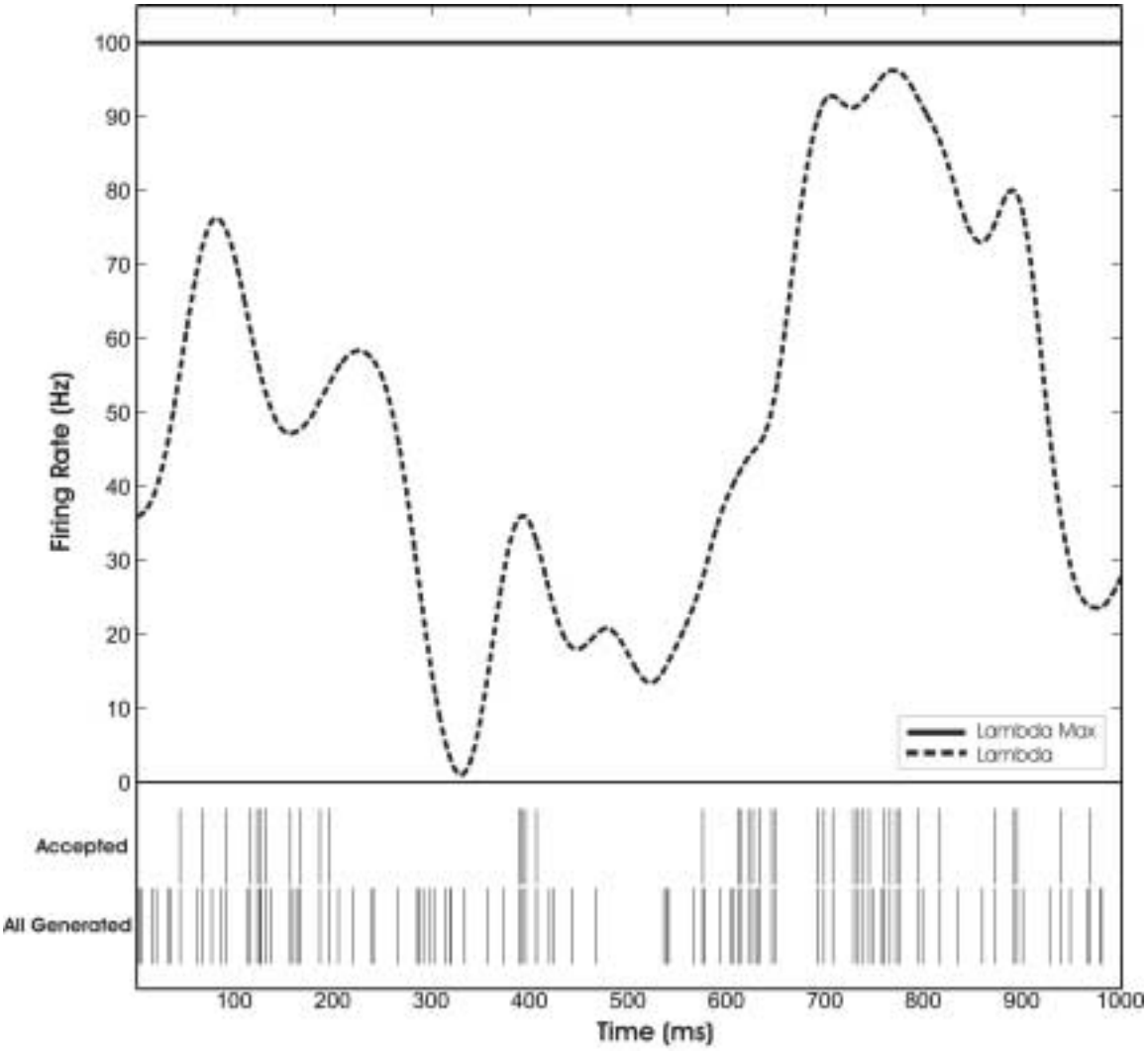

Figure 6. Illustration of the thinning algorithm. Initially, a set of candidate spikes (All Generated) are simulated from the simple Poisson process with parameter $\lambda_{\max}$ (—). The candidate spikes are thinned using the algorithm in Eq. to obtain a sample (Accepted) from the point process with conditional intensity function $\lambda(t \mid H_t)$ (----).