

Likelihood Methods for Point Processes with Refractoriness

Luca Citi, Demba Ba, Emery N. Brown, and Riccardo Barbieri

Dept. Anesthesia, Massachusetts General Hospital – Harvard Medical School, Boston, MA, USA.

Dept. Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA.

Keywords: Neural Engineering; Point processes; Encoding and decoding neural signals; Refractoriness; Refractory period; Brain-Machine Interfaces

Abstract

Likelihood-based encoding models founded on point-processes have received significant attention in the literature because of their ability to significantly improve neural decoding in Brain-Machine Interface applications. We propose an approximation to the likelihood of a point-process model of neurons which holds under assumptions about the continuous time process that are physiologically reasonable for neural spike trains: the presence of a refractory period, the predictability of the conditional intensity function, and its integrability. These are properties which apply to a large class of point-processes arising in applications other than neuroscience. The proposed approach has several advantages over state-of-the-art conventional ones. In particular, one can use

standard fitting procedures for generalized linear models (GLMs) based on iteratively-reweighted least-squares (IRWLS) while improving the accuracy of the approximation to the likelihood and reducing bias. As a result, the proposed approach can use a larger bin size to achieve the same accuracy as conventional approaches would with a smaller bin size. This is particularly important when analyzing neural data with high mean and instantaneous firing rates. We demonstrate these claims on simulated and real neural spiking activity. By allowing a substantive increase in the required bin size, our algorithm has the potential of lowering the barrier to the use of point-process methods in an increasing number of neural engineering applications.

1 Introduction

Technological advances of the past decade have given us an unprecedented glimpse into the inner workings of the brain. A common paradigm in experimental neuroscience is to record the activity of groups of neurons in various behavioural settings and/or in response to sensory stimuli. In computational neuroscience, this paradigm is exploited by building models of the encoding process from behaviour/stimulus to neural responses. In recent years, likelihood-based encoding models based on point-processes have received significant attention in the literature (Paninski et al., 2007; Barbieri et al., 2001; Wang and Principe, 2010; Berger et al., 2011; Meng et al., 2011; So et al., 2012) because of their ability to significantly improve neural decoding tasks (Brown et al., 2001; Paninski et al., 2010; Srinivasan et al., 2006; Barbieri et al., 2005b). The goal of encoding models is to improve our understanding of brain function and, most importantly,

to help to design algorithms for inferring behaviour/stimulus from previously unseen neural data, a process referred to as decoding and commonly used in the design of brain-machine interfaces (BMIs) (Shoham et al., 2005; Srinivasan et al., 2007; Sanchez et al., 2008).

Likelihood-based models of neural spike trains are based on a continuous-time point-process model of a neuron. A point-process is fully characterized by its conditional intensity function (CIF), which generalizes the notion of rate of a Poisson process to include time and history dependence. Generalized linear models (GLMs) are a class of discrete-time models based on log-linear expansions of a discrete-time version of the conditional intensity function (CIF) of a point process (Berman and Turner, 1992; Paninski, 2004; Truccolo et al., 2005). Berman and Turner (1992) were among the first to suggest the use of discretization for approximate maximum likelihood estimation of point-process models using the GLM framework. In neuroscience, this has resulted in the development of successful frameworks to characterize the dynamics of various neural systems (Paninski, 2004; Truccolo et al., 2005) and to develop algorithms for decoding based on neural data, with applications to BMIs (Chiel and Thomas, 2011).

In this paper, we demonstrate that this approximation can be significantly improved by imposing a biophysically-motivated structure on the continuous-time CIF of a point-process model of a neuron. Coincidentally, this translates into a specific, data-dependent, form of the weighting function in Berman's approximation. The proposed approximation holds under assumptions about the continuous time process that are reasonable for neural spike trains: the presence of a refractory period, the predictability and integrability of the CIF. These properties are not exclusive to neural point processes, but also

apply to point process models of geysers, heart beats and repeated failures of components in engineering systems (see Section 2.4).

The resulting point-process likelihood has several advantages over state-of-the-art approximations (Paninski, 2004; Truccolo et al., 2005). First, one can use standard fitting procedures for GLMs based on iteratively-reweighted least-squares (IRWLS) while improving the accuracy of the approximation to the likelihood and reducing bias. Stated otherwise, the proposed approach can use a larger bin size to achieve the same accuracy as the mentioned approaches with a smaller bin size. This latter aspect is particularly important when analyzing neurons with high spiking rates which often require a sampling frequency of at least 10 kHz. Tested on one example of this type of recordings, our method achieved at 1 kHz results comparable to those of a Poisson GLM (Paninski, 2004; Truccolo et al., 2005) at 10 kHz. This noteworthy improvement might translate into a reduced computational cost, reduced storage cost and less stringent requirements for the hardware acquisition and processing chain of BMIs. Second, we find that our formulation partly obviates the use of the correction techniques introduced in (Haslinger et al., 2010) for goodness-of-fit assessment using the time-rescaling theorem and discrete-time approximations to the CIF. We demonstrate our claims on simulated data, as well as real data from rat thalamic neurons recorded in response to periodic whisker deflections varying in velocity (Temereanca et al., 2008). These data are characterized by high mean and instantaneous firing rates, on the order of 20 and 200 Hz respectively.

The remainder of our treatment proceeds as follows. First, we introduce our model of continuous-time point processes with refractoriness (Section 2), along with the re-

sulting discrete-time likelihoods. In this section, we also compare the latter to conventional discrete-time approximations of continuous-time point process likelihoods. We use simulations to demonstrate the power of our new model which accounts for refractoriness. Then, in Section 3, we propose a class of parametric log-linear models and discuss the connections with conventional GLMs and iteratively re-weighted least-squares. We apply the log-linear models to analysis of real neural data. Finally, we provide concluding remarks in Section 4.

2 A model of continuous-time point process with refractoriness

We consider a continuous-time point process defined in $(0, T]$, with conditional intensity function (CIF) $\lambda(t | \mathcal{H}_t)$ where \mathcal{H}_t is a left-continuous filtration representing the history of the point process, of its afferent point process covariates, of other covariates and of deterministic inputs up to, but excluding, time t . Let $N(t)$ be the counting process associated with the point process. For each realization, $N(t)$ is a càdlàg (*continue à droite, limite à gauche*; right continuous with left limits) function counting the number of events observed up to, and including, time t (Andersen et al., 1995) (see Figure 1).

2.1 Key assumptions

We assume that the continuous-time point process belongs to the family of point processes for which the following properties hold.

(P1) Refractoriness: The CIF, $\lambda(t | \mathcal{H}_t)$, is a piecewise Lipschitz continuous function

in t which jumps to zero immediately after the occurrence of every event:

$$\lim_{\epsilon \rightarrow 0^+} \lambda(t + \epsilon | \mathcal{H}_{t+\epsilon}) = 0 \quad \forall t : dN(t) = 1 . \quad (1)$$

A graphical representation of this property is given in Figure 1. This assumption is plausible for point processes with a refractory period such as neural spike trains, motor-unit firing trains (Hu et al., 2004), and others (see Section 2.4).

(P2) Predictability: The CIF changes slowly in response to new information coming from covariates and deterministic inputs. For all $\epsilon > 0$ there exist $\Delta > 0$ such that, given $s \in (0, T]$, $\tau \in (0, \Delta]$ and $N(s) = N(s + \tau)$, we have:

$$\sup_{t \in (s, s+\tau)} |\lambda(t | \mathcal{H}_t) - \lambda(t | \mathcal{H}_s)| < \epsilon \lambda(t | \mathcal{H}_s) . \quad (2)$$

(P3) Integrability: Given the history at a time s , the function $\lambda(t | \mathcal{H}_s)$ is Riemann integrable for $t \in (s, T]$. This implies that for all $\epsilon > 0$ there exists $\Delta > 0$ such that for $t_1 \in (s, s + \Delta]$:

$$\left| \lambda(t_1 | \mathcal{H}_s) \Delta - \int_s^{s+\Delta} \lambda(t | \mathcal{H}_s) dt \right| < \epsilon . \quad (3)$$

Therefore, we can define a representative value $\tilde{\lambda}$ of $\lambda(t | \mathcal{H}_s)$ in $(s, s + \Delta]$ such that the former approximates the integral mean value of the latter inside the interval:

$$\tilde{\lambda} \simeq \frac{1}{\Delta} \int_s^{s+\Delta} \lambda(t | \mathcal{H}_s) dt . \quad (4)$$

[Figure 1 about here.]

2.2 Implications of key assumptions on discrete-time approximation

Given Δ , small enough so that the key assumptions hold and that the probability of two events in the same bin is negligible, we partition the observation interval in I bins of width Δ and attempt to relate the original continuous-time point process with the resulting discrete-time point-process.

Probabilities of key events: When (P2) and (P3) apply, the probability of observing no events in $(\bar{i}\Delta, i\Delta]$ (\bar{i} is short for $i-1$) can be approximated as:

$$\Pr(N(i\Delta) - N(\bar{i}\Delta) = 0) = e^{-\int_{\bar{i}\Delta}^{i\Delta} \lambda(t|\mathcal{H}_s) dt} \simeq e^{-\int_{\bar{i}\Delta}^{i\Delta} \lambda(t|\mathcal{H}_{\bar{i}\Delta}) dt} \simeq e^{-\lambda_i \Delta} \quad (5)$$

where λ_i is the representative value of the CIF inside the i -th bin according to (4). If (P1) also applies and Δ is sufficiently small, the probability of more than one event in a given bin is negligible because the first event causes the CIF to be approximately zero for the remaining of the bin. In fact, while for a generic point process the probability of two events occurring in the same bin of infinitesimal size Δ is $O(\Delta^2)$, when (P1) holds, one can prove that this probability is $O(\Delta^3)$ (see Appendix A). Therefore, the probability of observing exactly one event is simply $1 - e^{-\lambda_i \Delta}$. As expected, for Δ infinitesimal, the first order Taylor expansions of the probability of no events and of that of one event coincide with $1 - \lambda_i \Delta$ and $\lambda_i \Delta$, respectively, i.e., with the results of a

Bernoulli approximation of the Poisson process.

Discrete-time point-process PMF: We define the indicator function of the discrete-time point process as $\Delta N_i = \min\{1, N(i\Delta) - N(\bar{i}\Delta)\}$, a function that takes value zero if no events were observed in the i -th bin and one otherwise. Therefore, a given realization of the point process can be represented as a binary sequence $\Delta N_{1:I} = \{\Delta N_i\}_{i=1}^I$ with probability mass function (PMF):

$$P_n(\Delta N_{1:I}) = \prod_{i=1}^I (1 - e^{-\lambda_i \Delta})^{\Delta N_i} (e^{-\lambda_i \Delta})^{1 - \Delta N_i} \quad (6)$$

where we use the subscript “n” to remind us that this is the PMF of a discrete-time version of the continuous-time point process obtained using the properties of neural spike trains, in particular the presence of a refractory period, as per (P1). The log-probability is simply:

$$\ln P_n(\Delta N_{1:I}) = \sum_{i=1}^I \Delta N_i \ln(1 - e^{-\lambda_i \Delta}) - (1 - \Delta N_i) \lambda_i \Delta. \quad (7)$$

In Appendix B we prove that, as $\Delta \rightarrow 0$, (7) converges to the continuous time case:

$$\ln f(N_{0:T}) = \int_0^T \ln \lambda(t | \mathcal{H}_t) dN(t) - \int_0^T \lambda(t | \mathcal{H}_t) dt. \quad (8)$$

Useful approximation: We introduce an approximation of (7) that leads to a simpler and computationally attractive form for the log-likelihood, once a log-linear parametriza-

tion of λ_i is adopted. Noting that

$$\ln\left(\frac{1 - e^{-\xi}}{\xi}\right) = -\frac{\xi}{2} + o(\xi) \quad \text{as } \xi \rightarrow 0, \quad (9)$$

(7) can be approximated as

$$\begin{aligned} \ln P_n(\Delta N_{1:I}) &= \sum_{i=1}^I \Delta N_i \left[\ln(\lambda_i \Delta) + \ln\left(\frac{1 - e^{-\lambda_i \Delta}}{\lambda_i \Delta}\right) \right] - (1 - \Delta N_i) \lambda_i \Delta \\ &= \sum_{i=1}^I \Delta N_i \ln(\lambda_i \Delta) - \left(1 - \frac{\Delta N_i}{2}\right) \lambda_i \Delta + o(\Delta). \end{aligned} \quad (10)$$

Therefore, we define the approximated likelihood as:

$$\ln P_{n'}(\Delta N_{1:I}) = \sum_{i=1}^I \Delta N_i \ln(\lambda_i \Delta) - \left(1 - \frac{\Delta N_i}{2}\right) \lambda_i \Delta. \quad (11)$$

We will see in Section 2.5 that in most cases this approximation introduces an acceptable error. Later, in Section 3, we will show how this formulation allows an efficient procedure of fitting a log-linear model by means of an IRWLS algorithm. One can interpret (11) as a data-dependent choice of the weighting proposed by Berman and Turner (1992). We would like to stress that Berman's motivation behind the use of weighting functions is mostly heuristic. However, in our setting, the weighting function is dictated by the conditions imposed on the continuous-time point process, in particular refractoriness.

2.3 Comparison with conventional discrete-time approximation

We now compare (11) with a discretized version of (8) which is usually taken as (see Truccolo et al., 2005, Equation (3)):

$$\ln P_d(\Delta N_{1:I}) = \sum_{i=1}^I \Delta N_i \ln(\lambda_i \Delta) - \lambda_i \Delta . \quad (12)$$

While both approximations, (11) and (12), converge in the limit to the continuous PDF (8), they do so with a different linear coefficient for Δ . Specifically, while the first term is the same in both cases, the second term differs. From an empirical point of view, we can think of this second term as the numerical computation of the integral in the second term of (8) and the two implementations, (11) and (12), differ only in the way the CIF is integrated inside those bins containing an event. The contribution of one such bin to the second term of (8) is:

$$\int_{\bar{i}\Delta}^{i\Delta} \lambda(t | \mathcal{H}_t) dt \Big|_{\Delta N_i=1} \simeq \begin{cases} \lambda_i \Delta / 2 & \text{in (11),} \\ \lambda_i \Delta & \text{in (12).} \end{cases} \quad (13)$$

Defining $\tau = \sup\{0 < \tau \leq \Delta \mid N(\bar{i}\Delta + \tau) = N(\bar{i}\Delta)\}$, i.e., the time when the event occurs relative to the start of the bin, and using the properties (P1), (P2) and (P3) in this order, we obtain:

$$\int_{\bar{i}\Delta}^{i\Delta} \lambda(t | \mathcal{H}_t) dt \Big|_{\Delta N_i=1} \simeq \int_{\bar{i}\Delta}^{\bar{i}\Delta + \tau} \lambda(t | \mathcal{H}_t) dt \simeq \int_{\bar{i}\Delta}^{\bar{i}\Delta + \tau} \lambda(t | \mathcal{H}_{\bar{i}\Delta}) dt \simeq \lambda_i \tau . \quad (14)$$

We use a one-jump point process to find the expected value of τ conditioned to the fact that an event occurred inside the bin:

$$E[\tau | \Delta N_i = 1] = \frac{\int_0^{\Delta} \tau \lambda_i e^{-\lambda_i \tau} d\tau}{\int_0^{\Delta} \lambda_i e^{-\lambda_i \tau} d\tau} = \frac{e^{-\lambda_i \Delta} e^{\lambda_i \Delta} - \lambda_i \Delta - 1}{\lambda_i (1 - e^{-\lambda_i \Delta})} = \frac{1}{2} \Delta + o(\Delta) \quad (15)$$

Replacing (15) in (14) we see that for those bins where an event occurs, approximately half of $\lambda_i \Delta$ should be considered. Looking at (13), we see that this corresponds to what our solution does, while (12) tends to overestimate the LHS of (13).

2.4 Point-processes with refractoriness are pervasive

The assumptions about the continuous time process under which the approximation holds (the presence of a refractory period, the predictability of the CIF, and its integrability) are not exclusive to neural point processes, but also apply to a wide spectrum of point processes, such as models of geysers, heart beats, and repeated failures of components in engineering systems (e.g., light bulbs).

There are two main approaches to modelling point-process data. One can either model the CIF, or the probability density (assuming it exists) of the inter-event-intervals (IEIs). In some sense, these two approaches are equivalent because the following one-to-one transformation between CIFs and IEI PDFs holds:

$$\lambda(t | \mathcal{H}_t) = \frac{f(t | \mathcal{H}_t)}{\int_t^{\infty} f(\tau | \mathcal{H}_t) d\tau} . \quad (16)$$

In this paper, we have adopted the former approach. In light of the above transforma-

tion, it is not hard to show that the refractory property of the CIF in (1) holds if and only if the IEI distribution has a PDF that vanishes as its argument goes to zero. Therefore, the proposed framework for modeling point-processes with refractoriness is also applicable to a much broader range of problems. For example, the log-normal, inverse-Gaussian, Weibull and gamma (with shape parameter $k > 1$) distributions, which are commonly used in survival analysis (Kalbfleisch and Prentice, 2002), models of heart beats (Barbieri et al., 2005a), geyser and seismic data (Nishenko and Buland, 1987), satisfy the property that their respective PDFs vanish as their argument goes to zero. The corresponding CIFs must therefore satisfy the refractory property of (1), which is the backbone of the framework developed in this paper.

2.5 Simulations

In this section, we simulate three different examples of renewal continuous-time point processes to compare, as a function of the bin size Δ , the accuracy of estimates of the log-PDF (and consequently the log-likelihood) obtained using (7), (11), and (12).

Example 1: The first example is a homogeneous Poisson (P) process which, by definition, has constant CIF λ_P and exponentially distributed inter-spike-intervals (ISIs), as shown in Figure 2. For such a process, (P2) and (P3) hold while (P1) does not.

Example 2: The second case considers a Rayleigh (R) distribution for the ISIs. The corresponding CIF is $\lambda_R(z) = z/\sigma_R^2$, which increases linearly with the time since the last event, z . We chose this example because it has the simplest form of CIF that also complies with (P1). The Rayleigh distribution is a special case of the Weibull

distribution; it is skewed and bell-shaped (see Figure 2) and can be used to approximate the ISIs of neural spike trains (Lansky and Greenwood, 2005; Tyrcha, 2008).

Example 3: The last example we consider is a renewal process for which the ISIs follow an inverse Gaussian (IG) distribution. The IG distribution is particularly suited to model spike trains because it represents the distribution of the first threshold passage time of an integrate-and-fire neuron driven by a Wiener process with drift (Brown, 2005). Given the time since the last event, the CIF corresponding to IG-distributed ISIs with mean μ and shape parameter k , is

$$\lambda_{\text{IG}}(z) = \frac{\alpha_3(z) \varphi(\alpha_1(z) - \alpha_2(z))}{\Phi(\alpha_1(z) - \alpha_2(z)) - e^{2\alpha_1(z)\alpha_2(z)} \Phi(\alpha_1(z) + \alpha_2(z))} \quad (17)$$

where the vector $\alpha(z) = [-\sqrt{kz}/\mu, -\sqrt{k/z}, \sqrt{k/z^3}]$ while $\varphi(x)$ and $\Phi(x)$ are the probability density function and the cumulative distribution function of a standard Gaussian random variable.

[Figure 2 about here.]

Simulation parameters: For each of the three distributions above, we simulated one realization in $(0, T]$ with $T = 300$ s, assuming that the last event before the observation interval happened at $u_0 = 0$. Given the set of event times, $\{u_k\}_{k=1}^{N(T)}$, the continuous time log-PDF is simply:

$$\ln f_D(N_{0:T}) = \sum_{k=1}^{N(T)} \ln f_D(u_k - u_{k-1}) + \ln(1 - \mathfrak{F}_D(T - u_{N(T)})) \quad (18)$$

where $D \in \{P, R, IG\}$ and $f_D(z)$ and $\mathfrak{F}_D(z)$ are the PDF and CDF associated with the ISIs.

We binned each process for 31 values of the bin size, Δ , logarithmically spaced between $10^{-4.5}$ and $10^{-1.5}$, i.e., roughly 0.03 ms and 30 ms. Then, we evaluated the log-PMF using (7), (11) and (12) in turn, followed by subtraction of the correcting factor $(\ln(\Delta) \sum_{i=1}^I \Delta N_i)$ to obtain an estimate of the log-PDF from the log-PMF. For large bin sizes, realizations of all three processes (and P in particular) may have more than one event in some bins. To account for this, we also tested a variant of (12) where ΔN_i is replaced with a non-capped version $\Delta N_i^* = N(\bar{i}\Delta) - N(i\Delta)$.

Results: Results of these simulations are presented in the right panels of Figure 2. For the homogeneous Poisson process, all methods of estimating the log-PDF starting from a discrete-time process, fail miserably unless a very small bin size is used. The only exception is the one that allows for more than one event per bin. As this type of point process does not comply with (P1), this result was expected.

Both for the Rayleigh and the inverse Gaussian distributions, the two formulations presented in this paper, (7) and (11), show a very remarkable improvement compared with the legacy approach of (12). For example, for the Rayleigh distribution, the conventional approach requires a bin size of 0.1 ms to achieve a result comparable to the one that our method shows at 5 ms, while for the IG distribution it requires 0.3 ms to achieve the same result that our method shows at around 4 ms.

3 Log-linear models of discrete-time point-process data

In this section, we compare our formulation of the point-process PMF with the legacy approach in parametric, likelihood-based modelling of neural data. Parametric, likelihood-based, models of neural spike trains use the log-PMF of the data as the likelihood function. It is natural to expect that inference algorithms for these models would benefit from our novel formulation of the discretized point-process likelihood. In the next section we will show that this is indeed the case.

Parametric models of the log-likelihood in (12) have been successfully applied to the identification of neural systems (Paninski, 2004; Truccolo et al., 2005). Their main advantage is the ability to relate neuronal firing to extrinsic covariates such as stimuli or intrinsic ones which capture the internal dynamics of neurons. In log-linear models, one expresses $\ln(\lambda_i)$ as a linear function of the covariates (Truccolo et al., 2005). This turns (12) into a generalized linear model (GLM) with Poisson observations and log link (Fahrmeir and Tutz, 2001). Below, we show that a log-linear model of λ_i in (11) leads to a parametric model which can be fit as efficiently as conventional GLMs of discrete-time point-process data (Paninski, 2004; Truccolo et al., 2005).

3.1 IRWLS algorithm for log-linear model fitting

GLMs are a generalization of linear least-squares to observations from the exponential family. GLMs are computationally very attractive because they can be fit by iteratively-reweighted least-squares (IRWLS), that is to say by solving a sequence of weighted least-squares problems (Fahrmeir and Tutz, 2001).

If we assume $\ln(\lambda_i) = \beta^T x_i$, for a d -dimensional vector of parameters β and a

vector x_i of covariates of the same dimension, then (11) becomes

$$\ln P_{n'}(\Delta N_{1:I}, \beta; X) \propto \sum_{i=1}^I \Delta N_i \beta^T x_i - \left(1 - \frac{\Delta N_i}{2}\right) \exp(\beta^T x_i) \quad (19)$$

where X is the I -by- d matrix with x_i^T in the rows. One can maximize (19) using Newton's method. In order to draw a parallel between (19) and conventional GLMs, we consider the following generalization of (19):

$$\ln P_m(\Delta N_{1:I}, \beta; X, \rho) \propto \sum_{i=1}^I \Delta N_i \beta^T x_i - \rho_i \exp(\beta^T x_i) \quad (20)$$

where $\rho_i = 1 - \frac{\Delta N_i}{2}$ and $m = n'$ for (19), while $\rho_i = 1$ and $m = d$ for conventional GLMs arising from maximizing P_d in (12) using a log-linear parametrization of λ_i . In Appendix C, we derive the IRWLS algorithm for maximizing (20) showing that, up to a choice of a weighting function (ρ), maximizing (19) and fitting conventional GLMs are equivalent. This implies that fast implementations (Komarek and Moore, 2005) of IRWLS can be modified in a simple way to maximize the log-likelihood (19).

3.2 Goodness-of-fit assessment

Quantitatively measuring the agreement between a proposed model and a spike train is crucial for establishing the model's validity prior to using it to make inferences about the neural system being studied. The time rescaling theorem can be used to check the goodness of fit of statistical models of neural spike trains (Brown et al., 2002).

In a recent study, Haslinger and colleagues (2010) have drawn attention to the fact that the finite temporal resolution of discrete time models introduces some bias in the

application of the time rescaling theorem, which can lead to misleading results, that is, indicating poor goodness of fit even for accurate models. They found two root causes and proposed analytical corrections for these effects. The first cause is that, as an unavoidable consequence of binning, there is a lower bound on the smallest possible ISI (one bin). The second arises using Bernoulli discrete models (with logit link) and then naïvely applying the time rescaling theorem assuming $p_i = \lambda_i \Delta$ as the probability of a spike in a given bin.

Our method does not suffer from the second issue because it directly estimates λ_i , a discretized version of the continuous time function $\lambda(t)$, rather than the probability of one event in a discrete bin (as in the Bernoulli case). As we have presented in Section 2.2, when properties (P1)–(P3) hold, these two variables are related by $p_i = \Pr(\Delta N_i = 1) = 1 - e^{-\lambda_i \Delta}$ which is exactly the inverse of the correction that the authors propose in equation (2.35) of their manuscript.

In our proposed approach, the goodness-of-fit assessment can be simply performed using the following procedure, adapted from Haslinger et al. (2010):

1. Use the loglinear model (20), based on (11) or (12), to fit a model to the observed spike train and obtain an estimate of the sampled CIF, $\hat{\lambda}_i$;
2. Given the set of bins containing a spike $\{i_k \mid \Delta N_{i_k} = 1\}$, for each ISI find the integrated CIF as:

$$\xi_k = \sum_{i=i_k+1}^{i_{k+1}-1} \hat{\lambda}_i \Delta + \hat{\lambda}_{i_{k+1}} \tau_k \quad (21)$$

where τ_k represents the random time of the event relative to the start of the bin, and can be obtained by first drawing a random variable r_k uniformly distributed

in $[0, 1]$ and then calculating

$$\tau_k = -\frac{\ln[1 - r_k (1 - \exp(-\hat{\lambda}_{i_{k+1}}\Delta))]}{\hat{\lambda}_{i_{k+1}}}. \quad (22)$$

In most cases this expression can be approximated as $\tau_k = r_k\Delta$.

3. Finally, after a further variable transform, $z_k = 1 - e^{-\xi_k}$, the goodness-of-fit can be assessed by plotting the empirical CDF of the transformed ISIs, z_k , against the CDF of the uniform distribution.

3.3 Simulations

In this section we present the results of the comparison between the conventional GLM approach and our novel log-linear model (19), in the estimation of the regression terms of $\ln \lambda(t)$. We assumed a simple, yet plausible, form for $\ln \lambda(t)$ allowing for simulation in continuous time and then assessed the estimation accuracy as a function of bin size using either method.

Generation of simulated spike trains: We modeled $\ln \lambda(t)$ as the convolution of an autoregressive kernel $\psi(z)$ with the comb of previously observed events:

$$\ln \lambda(t) = \ln(\ell_0) + \int_0^t \psi(t - \tau) dN(\tau) \quad (23)$$

where ℓ_0 is a constant term. We used the kernel function

$$\psi(z) = \begin{cases} \ln \left(c_1 \left(\frac{z}{c_3} \right)^3 + c_2 \left(\frac{z}{c_3} \right)^2 \right) & \text{if } 0 < z < c_3, \\ 0 & \text{otherwise,} \end{cases} \quad (24)$$

and set the vector $c = [-8, 9, 0.1]$ and $\ell_0 = 100$, see Figure 3. We chose this specific kernel function because it allows for the simulation of spike trains in continuous time without resorting to discretization and because it results in a physiologically-plausible distribution of the ISIs, as confirmed by Figure 4. The CIF and its integral, $\Lambda(t)$, are piecewise polynomials in t and, from (23) and (24), we obtain

$$\lambda(t) = \ell_0 \prod_{k=N(t-c_3)+1}^{N(t)} \left(c_1 \left(\frac{t - u_k}{c_3} \right)^3 + c_2 \left(\frac{t - u_k}{c_3} \right)^2 \right) \quad (25)$$

where $\{u_k\}_{k=1}^{N(T)}$ is the ordered set of event times. To simulate a spike train, the time rescaling theorem (Brown et al., 2002) was used to calculate the next event u_{k+1} from the set of previous events as the solution of $\Lambda(u_{k+1}) - \Lambda(u_k) = -\ln(q_{k+1})$ where q_{k+1} is a random variable uniformly sampled in $(0, 1)$. The exact solution was found using a polynomial root finding algorithm as one of the eigenvalues of the companion matrix (Edelman and Murakami, 1995). Note that the CIF of Figure 1 was generated in this fashion.

[Figure 3 about here.]

Using this procedure, we generated 250 realizations of the continuous-time point process, each one with an observation interval of 600 s and approximately 40 events

per second. The resulting distribution of ISIs is reported in Figure 4.

[Figure 4 about here.]

Model fitting: We considered the following model for the discrete-time CIF:

$$\ln \lambda_i = \beta_0 + \sum_{j=1}^{\lceil c_3/\Delta \rceil} \beta_j \Delta N_{i-j} \quad (26)$$

which is the discrete-time equivalent of (23). Then, we fitted the conventional GLM and our log-linear model on each realization for 9 values of Δ , logarithmically spaced between $10^{-3.5}$ and $10^{-1.5}$, that is, approximately between 0.3 ms and 30 ms.

Results: We assessed the difference between the original baseline firing rate $\ell_0=100$ and the value $\ln \beta_0$ estimated by the two methods, as a function of Δ . The results are presented in Figure 5. Compared to the conventional GLM, our method approaches the ideal value of ℓ_0 with higher accuracy and less sensitivity to the bin size. In fact, the bin size can be increased without much detriment up to the value (approximately 0.1 s) where the probability of more spikes in a bin becomes significant, while the accuracy of the traditional GLM progressively decreases for all bin sizes. We also compared the accuracy of the two methods at the task of recovering the autoregressive kernel. Figures 6(a)–6(c) show the true kernel $\psi(z)$ and the estimated kernel β_j for three values of Δ , while Figure 6(d) shows the error $\beta_j - \psi(z)$, with $j = \lfloor z/\Delta + 1/2 \rfloor$, for $z = 0.02$ s and $z = 0.07$ s as a function of Δ . From these results, we notice that the traditional GLM, which maximizes the likelihood given by (12), tends to underestimate the base firing rate ℓ_0 , overestimate the kernel for small values of z (i.e., briefly after each event) and

underestimate it for bigger z . Our log-linear model, which maximizes (7), instead, is able to estimate the kernel with much higher accuracy.

[Figure 5 about here.]

[Figure 6 about here.]

3.4 Real data

In this section, we demonstrate that the improvement that our method shows on simulated data also holds for real neural spike trains. In particular, we tested our approach on single unit recordings from individual thalamic barreloid neurons in the somatosensory whisker/barrel system of an anaesthetized rat.

Experiment: Here, we provide a brief description of the experimental protocol that was reported in detail by Temereanca et al. (2008). The principal whisker was deflected in caudal direction by means of a piezoelectric stimulator using periodic waveforms of different velocity delivered at a repetition rate of eight per second. Each deflection was 1 mm in amplitude and began and ended from the whiskers neutral position as the trough (phase of $3\pi/2$) of a single sine wave. The fast stimulus was generated by taking one period of a 40 Hz sine wave and delivering it every 125 ms for 2 s. The other two stimuli were generated similarly, by taking one cycle of a 25 Hz sine wave and one cycle of an 8 Hz sine wave. The whisker was deflected pseudorandomly using the three stimuli, and pseudorandomized blocks were delivered 50 times, with interstimulus intervals of 1.5 s. Spike waveforms were analyzed off-line, clustered based on waveform shape and inter-spike interval, and the spike stamps saved at 100 μ s resolution.

Log-linear model of the CIF: In our analysis, we assumed a log-linear CIF given by:

$$\ln \lambda_i = \beta_0 + \gamma_1 s(i\Delta) + \gamma_2 s(i\Delta - q_1) + \sum_{j=1}^{q_2/\Delta} \beta_j \Delta N_{i-j} \quad (27)$$

where $s(t)$ is the stimulus, i.e., the whisker deflection at time t , while $q_1 = 1$ ms and $q_2 = 6$ ms. These parameter values were chosen based on prior extensive analyses of the neurons considered here using state-of-the-art methods (Paninski, 2004; Truccolo et al., 2005). We used the conventional GLM and our log-linear model to estimate the vector $\beta = [\beta_0, \dots, \beta_{q_2/\Delta}, \gamma_1, \gamma_2]$ that best explains the observed spike train given the stimulus and the process history. We repeated the analysis for bin sizes $\Delta \in \{0.1, 0.2, 0.5, 1\}$ milliseconds.

Results: The top three panels of Figure 7 show the results of the estimation of β_0 , γ_1 , and γ_2 as a function of the bin size when using conventional GLM and our log-linear model. Although in the case of real data the ground truth is unknown, it is fair to assume that the true values are more likely to be somewhere around those obtained with a finer sampling. It is clear from the figure that, when using the traditional GLM based on (12), there is a strong effect of the bin size on the estimated values, leading to a significant bias when a coarser sampling is used. This effect is much less pronounced when using the log-linear model coming from our new likelihood formulation (11). In fact, the values obtained with our model at 1 ms (1 kHz sampling rate) are similar to those obtained by the GLM for 0.1 ms (10 kHz).

Using the time-rescaling procedure described in Section 3.2, we performed a goodness-of-fit analysis and assessed the Kolmogorov-Smirnov (KS)-distance achieved by the

fitted models for different bin sizes. The results are reported in the bottom panel of Figure 7, which confirms the increasing inaccuracy of the GLM model for increasing bin sizes. Our log-linear model, instead, achieves approximately the same fit for all the sampling rates tested. For even larger bin sizes (not shown in the figure), the fit degrades rapidly for both methods as soon as the assumption of one spike per bin is violated, similarly to what is shown in Figure 2 for the likelihood.

[Figure 7 about here.]

4 Discussion

The traditional discretized version of the Poisson process PDF (12), as reported by Truccolo et al. (2005), serves as a building block for many advanced statistical analysis and signal processing techniques (Paninski et al., 2007; Friedman et al., 2010; Zhao et al., 2011; Zhao and Iyengar, 2010; Pillow et al., 2011; Paninski et al., 2010; Czanner et al., 2008; Lawhern et al., 2010) applied to neural signals. These techniques have been used for basic neuroscience research (Kass et al., 2011; Okatan et al., 2005; Eldawlatly et al., 2009; Berger et al., 2011; Jenison et al., 2011; So et al., 2012), to improve biophysical neural models (Ahrens et al., 2008; Meng et al., 2011; Mensi et al., 2012), or to design better BMIs (Shoham et al., 2005; Srinivasan et al., 2006, 2007; Truccolo et al., 2008; Wang and Principe, 2010; Saleh et al., 2012).

In this paper we presented a new formulation for the probability mass function of observing discrete-time realizations of continuous time point processes arising from neural spike trains. This new theory holds under assumptions about the continuous

time point process that are reasonable for neural spike trains: the presence of a refractory period, the predictability of the conditional intensity function, and its integrability within a time bin. These properties are not exclusive to neural point processes, but also apply to a much wider spectrum of point processes, including models of geysers, heart beats, and repeated failures of components in engineering systems. Our new definition represents a remarkable theoretical improvement over the traditional discretized version of the continuous time point-process PDF, in presence of refractoriness. As a result, the estimated value approaches the solution of the continuous problem, as the bin size goes to zero, at significantly higher speed of convergence.

Based on this new definition, we also introduced a log-linear model which shows an improvement over a traditional GLM fit. Both in simulations and with real data, our novel algorithm converges to the asymptotic value for bin sizes one order of magnitude larger than the traditional GLM. This can be advantageous when analyzing neurons with high firing rate. For example, we showed that our method achieves, on neural data resampled at 1 KHz, more refined outcomes (e.g., in terms of goodness of fit) than those of a Poisson GLM on the original recordings sampled at 10 kHz. This noteworthy improvement might help reduce computational and storage costs, as well as reduce the strain on the hardware acquisition and processing chain of BMIs. The improvement shown by our log-linear model over the traditional GLM is likely to extend to most models and algorithms based on the point process GLM framework, and comes at virtually no cost because, as we have shown with our log-linear model, the number of floating point operations is practically the same. In most cases, instead, the computational cost can decrease drastically because our new definition allows for the use of a

coarser sampling rate while still providing a comparable accuracy.

In conclusion, our novel likelihood formulation for point processes with refractoriness improves over conventional approaches that ignore the refractoriness of neural point process. As it only requires a minor modification to the likelihood term, it can replace the legacy point process likelihood (Truccolo et al., 2005) in virtually all instances where such a probabilistic framework is used. By allowing a substantive increase in the required bin size, our algorithm has the potential of lowering the barrier to the use of point-process methods in an increasing number of neural engineering applications.

Acknowledgments

The authors would like to thank Simona Temereanca for kindly providing the experimental data used in Section 3.4.

This work was supported by National Institutes of Health (NIH) through grants R01-HL084502 (R.B.) and DP1-OD003646 (E.N.B.).

A Asymptotic probability of more than one event in a bin

A function $f(t)$ is piecewise Lipschitz continuous in an interval $(a, b]$ if there is a positive constant C and a partition $\bigcup_{j=1}^J I_j = (a, b]$ such that

$$|f(t_1) - f(t_2)| \leq C |t_1 - t_2| \quad \forall t_1, t_2 \in I_j, 1 \leq j \leq J. \quad (28)$$

In our specific case, using the Lipschitz condition and the fact that the only jumps allowed are decreasing (the jumps to zero after each event), we obtain that the following condition holds for the CIF:

$$0 \leq \lambda(t_2 | \mathcal{H}_{t_2}) \leq \lambda(t_1 | \mathcal{H}_{t_1}) + C(t_2 - t_1) \quad \forall t_1, t_2 \in (0, T] : t_1 \leq t_2. \quad (29)$$

Let us consider the i -th bin and attempt to find the asymptotic behaviour of the probability that the number of events contained, $n = N(i\Delta) - N(\bar{i}\Delta)$, is at least one. Using the condition (29), we have that for $t \in (\bar{i}\Delta, i\Delta]$:

$$\lambda(t | \mathcal{H}_t) \leq \lambda(\bar{i}\Delta | \mathcal{H}_{\bar{i}\Delta}) + C(t - \bar{i}\Delta) \leq \lambda(\bar{i}\Delta | \mathcal{H}_{\bar{i}\Delta}) + C\Delta, \quad (30)$$

leading to:

$$\begin{aligned} \Pr(n \geq 1) &= 1 - \exp \left\{ - \int_{\bar{i}\Delta}^{i\Delta} \lambda(t | \mathcal{H}_t) dt \right\} \\ &\leq 1 - \exp \{ - [\lambda(\bar{i}\Delta | \mathcal{H}_{\bar{i}\Delta}) + C\Delta] \Delta \} = O(\Delta). \end{aligned} \quad (31)$$

Let us now find the probability of more than one event conditioned to the presence of at least one event in the bin, i.e., $\Pr(n > 1 | n \geq 1)$. Calling $t_1 \in (\bar{i}\Delta, i\Delta]$ the time when the first event occurs and using the property (P1), we have that $\lim_{\epsilon \rightarrow 0^+} \lambda(t_1 + \epsilon | \mathcal{H}_{t_1 + \epsilon}) = 0$ and therefore $\lambda(t | \mathcal{H}_t) \leq C(i\Delta - t_1) \leq C\Delta$ for $t \in (t_1, i\Delta]$. Observing that $n > 1$ implies at least one event in $(t_1, i\Delta]$, we can follow a reasoning similar to (31) and obtain:

$$\Pr(n > 1 | n \geq 1) \leq [1 - \exp(-C\Delta^2)] = O(\Delta^2). \quad (32)$$

Finally:

$$\Pr(n > 1) = \Pr(n > 1 | n \geq 1) \Pr(n \geq 1) = O(\Delta^3) \quad (33)$$

which is what we wanted to prove.

B Convergence for infinitesimal bin size

We can define the logarithm of the normalized PMF (log-nPMF) as

$$\begin{aligned} \ln \left(P_n(\Delta N_{1:T}) / \Delta^{N(T)} \right) &= -N(T) \ln \Delta \\ &+ \sum_{i=1}^I \Delta N_i \ln (1 - e^{-\lambda_i \Delta}) - (1 - \Delta N_i) \lambda_i \Delta \end{aligned} \quad (34)$$

and show that, in the limit $\Delta \rightarrow 0$, it converges to the logarithm of the continuous time probability density function (log-PDF) of observing exactly those $N(T)$ events in $(0, T]$, given in (8). We start by noting that, as $\Delta \rightarrow 0$:

$$\ln (1 - e^{-\lambda_i \Delta}) = \ln \left(\lambda_i \Delta \frac{1 - e^{-\lambda_i \Delta}}{\lambda_i \Delta} \right) = \ln \lambda_i + \ln \Delta + o(1) . \quad (35)$$

Then we take the limit of (34), replace (35), expand the terms, and finally obtain:

$$\begin{aligned} \lim_{\Delta \rightarrow 0} \ln \left(P_n(\Delta N_{1:T}) / \Delta^{N(T)} \right) &= -N(T) \ln \Delta \\ &+ \lim_{\Delta \rightarrow 0} \left(\sum_{i=1}^{\lceil T/\Delta \rceil} \Delta N_i \ln \lambda_i + \sum_{i=1}^{\lceil T/\Delta \rceil} \Delta N_i \ln \Delta - \sum_{i=1}^{\lceil T/\Delta \rceil} \lambda_i \Delta + \sum_{i=1}^{\lceil T/\Delta \rceil} \Delta N_i \lambda_i \Delta \right) \\ &= -N(T) \ln \Delta + \int_0^T \ln \lambda(t | \mathcal{H}_t) dN(t) + N(T) \ln \Delta - \int_0^T \lambda(t | \mathcal{H}_t) dt + 0 \end{aligned} \quad (36)$$

which simplifies to the continuous time log-PDF in (8).

C Log-linear model

Differentiating (20) with respect to β , we obtain its gradient and its Hessian:

$$\begin{aligned} g(\beta) &= \sum_{i=1}^I (\Delta N_i - \rho_i \lambda_i(\beta)) x_i = \sum_{i=1}^I (\Delta N_i - \rho_i \exp\{\beta^T x_i\}) x_i \\ &= X^T(\Delta N - \lambda_\rho(\beta)), \end{aligned} \quad (37)$$

$$H(\beta) = -\sum_{i=1}^I \rho_i \lambda_i(\beta) \cdot x_i x_i^T = -\sum_{i=1}^I \rho_i \exp\{\beta^T x_i\} \cdot x_i x_i^T = X^T W_\rho(\beta) X, \quad (38)$$

where X is the I -by- d matrix with x_i^T in the rows, ΔN and $\lambda_\rho(\beta)$ are I -length vectors satisfying with entries ΔN_i and $\rho_i \lambda_i(\beta)$ respectively, and $W_\rho(\beta)$ is the I -by- I diagonal matrix with diagonal elements $\rho_i \lambda_i(\beta)$.

One can maximize (20) by taking Newton steps as follows:

$$\beta_{(k+1)} = \beta_{(k)} - H^{-1}(\beta_{(k)})g(\beta_{(k)}). \quad (39)$$

Substituting (37) and (38) in (39) and re-arranging, we find that $\beta_{(k+1)}$ is the solution of a quadratic approximation to the objective function, which we refer to as weighted least-squares (WLS):

$$\beta_{(k+1)} = \operatorname{argmax}_{\beta} -\frac{1}{2}(b - A\beta)^T C(b - A\beta), \quad (40)$$

where $b = X\beta_{(k)} + W_\rho^{-1}(\beta_{(k)})(\Delta N - \lambda_\rho(\beta_{(k)}))$, $C = W_\rho(\beta_{(k)})$ and $A = X$. One

maximizes (20) by iteratively solving (40), hence the name iteratively re-weighted least-squares (IRWLS). Assuming X is full rank, (20) is a concave function of β . Therefore, there exists a unique solution to which the Newton algorithm, implemented by IRWLS, converges.

The general formulation (20) shows that the maximizing (19) and fitting conventional GLMs of neural data (using (12)) are equivalent up to the choice of ρ_i .

References

- M. B. Ahrens, L. Paninski, and M. Sahani. Inferring input nonlinearities in neural encoding models. *Network*, 19(1):35–67, 2008. doi:10.1080/09548980701813936.
- P. Andersen, O. Borgan, R. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer, New York, 1995.
- R. Barbieri, M. C. Quirk, L. M. Frank, M. A. Wilson, and E. N. Brown. Construction and analysis of non-poisson stimulus-response models of neural spiking activity. *Journal of Neuroscience Methods*, 105(1):25–37, Jan 2001. doi:10.1016/S0165-0270(00)00344-7.
- R. Barbieri, E. Matten, A. Alabi, and E. Brown. A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability. *American Journal of Physiology - Heart C*, 288(1):H424–H435, 2005a. doi:10.1152/ajpheart.00482.2003.
- R. Barbieri, M. A. Wilson, L. M. Frank, and E. N. Brown. An analysis of hippocam-

- pal spatio-temporal representations using a bayesian algorithm for neural spike train decoding. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(2):131–136, Jun 2005b. doi:10.1109/TNSRE.2005.847368.
- T. W. Berger, R. E. Hampson, D. Song, A. Goonawardena, V. Z. Marmarelis, and S. A. Deadwyler. A cortical neural prosthesis for restoring and enhancing memory. *Journal of Neural Engineering*, 8(4):046017, 2011. doi:10.1088/1741-2560/8/4/046017.
- M. Berman and T. Turner. Approximating point process likelihoods with GLIM. *Applied Statistics*, pages 31–38, 1992.
- E. Brown. *Theory of point processes for neural systems*, chapter 14, pages 691–726. *Methods and Models in Neurophysics*. Elsevier, 2005.
- E. Brown, R. Barbieri, V. Ventura, R. Kass, and L. Frank. The time-rescaling theorem and its application to neural spike train data analysis. *Neural Computation*, 14(2):325–346, 2002. doi:10.1162/08997660252741149.
- E. N. Brown, D. P. Nguyen, L. M. Frank, M. A. Wilson, and V. Solo. An analysis of neural receptive field plasticity by point process adaptive filtering. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21):12261–12266, Oct 2001. doi:10.1073/pnas.201409398.
- H. J. Chiel and P. J. Thomas. Applied neurodynamics: from neural dynamics to neural engineering. *Journal of Neural Engineering*, 8(6):060201, Dec. 2011. doi:10.1088/1741-2552/8/6/060201.
- G. Czanner, U. T. Eden, S. Wirth, M. Yanike, W. A. Suzuki, and E. N. Brown. Analysis

- of between-trial and within-trial neural spiking dynamics. *Journal of Neurophysiology*, 99(5):2672–2693, May 2008. doi:10.1152/jn.00343.2007.
- A. Edelman and H. Murakami. Polynomial roots from companion matrix eigenvalues. *Mathematics of Computation*, 64(210):763–776, 1995.
- S. Eldawlatly, R. Jin, and K. G. Oweiss. Identifying functional connectivity in large-scale neural ensemble recordings: a multiscale data mining approach. *Neural Computation*, 21(2):450–477, Feb 2009. doi:10.1162/neco.2008.09-07-606.
- L. Fahrmeir and G. Tutz. *Multivariate statistical modelling based on generalized linear models*. Springer Verlag, 2001.
- J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.
- R. Haslinger, G. Pipa, and E. Brown. Discrete time rescaling theorem: Determining goodness of fit for discrete time statistical models of neural spiking. *Neural Computation*, 22(10):2477–2506, 2010. doi:10.1162/NECO_a_00015.
- X. L. Hu, K. Y. Tong, and L. K. Hung. Oscillations in the power spectra of motor unit signals caused by refractoriness variations. *Journal of Neural Engineering*, 1(3):174–185, Sept. 2004. doi:10.1088/1741-2560/1/3/007.
- R. L. Jenison, A. Rangel, H. Oya, H. Kawasaki, and M. A. Howard. Value encoding in single neurons in the human amygdala during decision making. *Journal of Neuroscience*, 31(1):331–338, Jan 2011. doi:10.1523/JNEUROSCI.4461-10.2011.

- J. Kalbfleisch and R. Prentice. *The statistical analysis of failure time data*. Wiley, New York, 2nd edn edition, 2002.
- R. E. Kass, R. C. Kelly, and W.-L. Loh. Assessment of synchrony in multiple neural spike trains using loglinear point process models. *Annals of Applied Statistics*, 5(2B): 1262–1292, Jun 2011. doi:10.1214/10-AOAS429.
- P. Komarek and A. Moore. Making logistic regression a core data mining tool with TR-IRLS. In *Fifth IEEE International Conference on Data Mining*, pages 4–pp. IEEE, 2005. doi:10.1109/ICDM.2005.90.
- P. Lansky and P. E. Greenwood. Optimal signal estimation in neuronal models. *Neural Computation*, 17(10):2240–2257, Oct. 2005. doi:10.1162/0899766054615653.
- V. Lawhern, W. Wu, N. Hatsopoulos, and L. Paninski. Population decoding of motor cortical activity using a generalized linear model with hidden states. *Journal of Neuroscience Methods*, 189(2):267–280, Jun 2010. doi:10.1016/j.jneumeth.2010.03.024.
- L. Meng, M. A. Kramer, and U. T. Eden. A sequential Monte Carlo approach to estimate biophysical neural models from spikes. *Journal of Neural Engineering*, 8(6):065006, Dec 2011. doi:10.1088/1741-2560/8/6/065006.
- S. Mensi, R. Naud, C. Pozzorini, M. Avermann, C. C. H. Petersen, and W. Gerstner. Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms. *Journal of Neurophysiology*, 107(6):1756–1775, Mar 2012. doi:10.1152/jn.00408.2011.
- S. Nishenko and R. Buland. A generic recurrence interval distribution for earthquake

- forecasting. *Bulletin of the Seismological Society of America*, 77(4):1382–1399, 1987.
- M. Okatan, M. A. Wilson, and E. N. Brown. Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Computation*, 17(9):1927–1961, Sep 2005. doi:10.1162/0899766054322973.
- L. Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4):243–262, 2004.
- L. Paninski, J. Pillow, and J. Lewi. Statistical models for neural encoding, decoding, and optimal stimulus design. *Progress in Brain Research*, 165:493–507, 2007. doi:10.1016/S0079-6123(06)65031-0.
- L. Paninski, Y. Ahmadian, D. G. Ferreira, S. Koyama, K. R. Rad, M. Vidne, J. Vogelstein, and W. Wu. A new look at state-space models for neural data. *Journal of Computational Neuroscience*, 29(1-2):107–126, Aug 2010. doi:10.1007/s10827-009-0179-x.
- J. W. Pillow, Y. Ahmadian, and L. Paninski. Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains. *Neural Computation*, 23(1):1–45, Jan 2011. doi:10.1162/NECO_a_00058.
- M. Saleh, K. Takahashi, and N. G. Hatsopoulos. Encoding of coordinated reach and grasp trajectories in primary motor cortex. *Journal of Neuroscience*, 32(4):1220–1232, Jan 2012. doi:10.1523/JNEUROSCI.2438-11.2012.
- J. Sanchez, J. Principe, T. Nishida, R. Bashirullah, J. Harris, and J. Fortes. Technology

- and signal processing for brain-machine interfaces. *IEEE Signal Processing Magazine*, 25(1):29–40, 2008. doi:10.1109/MSP.2008.4408440.
- S. Shoham, L. M. Paninski, M. R. Fellows, N. G. Hatsopoulos, J. P. Donoghue, and R. A. Normann. Statistical encoding model for a primary motor cortical brain-machine interface. *IEEE Transactions on Biomedical Engineering*, 52(7):1312–1322, July 2005. doi:10.1109/TBME.2005.847542.
- K. So, A. C. Koralek, K. Ganguly, M. C. Gastpar, and J. M. Carmena. Assessing functional connectivity of neural ensembles using directed information. *Journal of Neural Engineering*, 9(2):026004, Apr 2012. doi:10.1088/1741-2560/9/2/026004.
- L. Srinivasan, U. T. Eden, A. S. Willsky, and E. N. Brown. A state-space analysis for reconstruction of goal-directed movements using neural signals. *Neural Computation*, 18(10):2465–2494, Oct 2006. doi:10.1162/neco.2006.18.10.2465.
- L. Srinivasan, U. T. Eden, S. K. Mitter, and E. N. Brown. General-purpose filter design for neural prosthetic devices. *Journal of Neurophysiology*, 98(4):2456–2475, Oct 2007. doi:10.1152/jn.01118.2006.
- S. Temereanca, E. N. Brown, and D. J. Simons. Rapid changes in thalamic firing synchrony during repetitive whisker stimulation. *Journal of Neuroscience*, 28(44):11153–11164, 2008. doi:10.1523/JNEUROSCI.1586-08.2008.
- W. Truccolo, U. Eden, M. Fellows, J. Donoghue, and E. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensem-

- ble, and extrinsic covariate effects. *Journal of Neurophysiology*, 93(2):1074, 2005. doi:10.1152/jn.00697.2004.
- W. Truccolo, G. M. Friehs, J. P. Donoghue, and L. R. Hochberg. Primary motor cortex tuning to intended movement kinematics in humans with tetraplegia. *Journal of Neuroscience*, 28(5):1163–1178, Jan 2008. doi:10.1523/JNEUROSCI.4415-07.2008.
- J. Tyrcha. Dynamics of integrate-and-fire models. *Mathematical Modeling of Biological Systems, Volume II*, pages 225–236, 2008. doi:10.1007/978-0-8176-4556-4_20.
- Y. Wang and J. C. Principe. Instantaneous estimation of motor cortical neural encoding for online brain-machine interfaces. *Journal of Neural Engineering*, 7(5):056010, Oct 2010. doi:10.1088/1741-2560/7/5/056010.
- M. Zhao and S. Iyengar. Nonconvergence in logistic and poisson models for neural spiking. *Neural Computation*, 22(5):1231–1244, May 2010. doi:10.1162/neco.2010.03-09-982.
- M. Zhao, A. Batista, J. P. Cunningham, C. Chestek, Z. Rivera-Alvidrez, R. Kalmar, S. Ryu, K. Shenoy, and S. Iyengar. An L1-regularized logistic model for detecting short-term neuronal interactions. *Journal of Computational Neuroscience*, Oct 2011. doi:10.1007/s10827-011-0365-5.

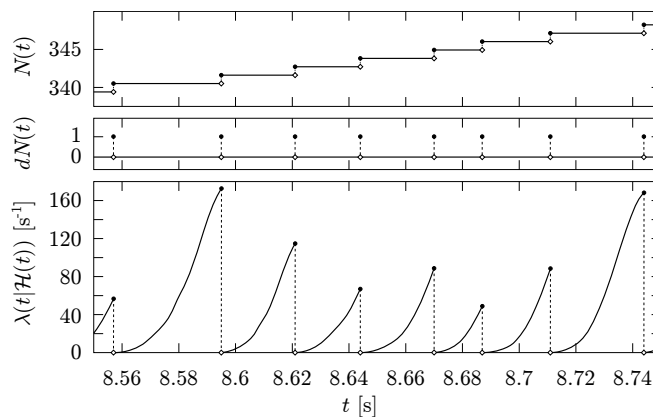


Figure 1: Representative example of a realization of a neural point process. The top plot shows the counting process, $N(t)$, which is a function counting the number of events observed up to, and including, time t . The central plot shows its differential, $dN(t)$, which is an indicator function that assumes value one if there is an event at time t and zero otherwise. The bottom plot shows the conditional intensity function, $\lambda(t | \mathcal{H}_s)$, which jumps to zero after every event because, as a result of the refractory period, an additional event cannot take place arbitrarily close to the previous one.

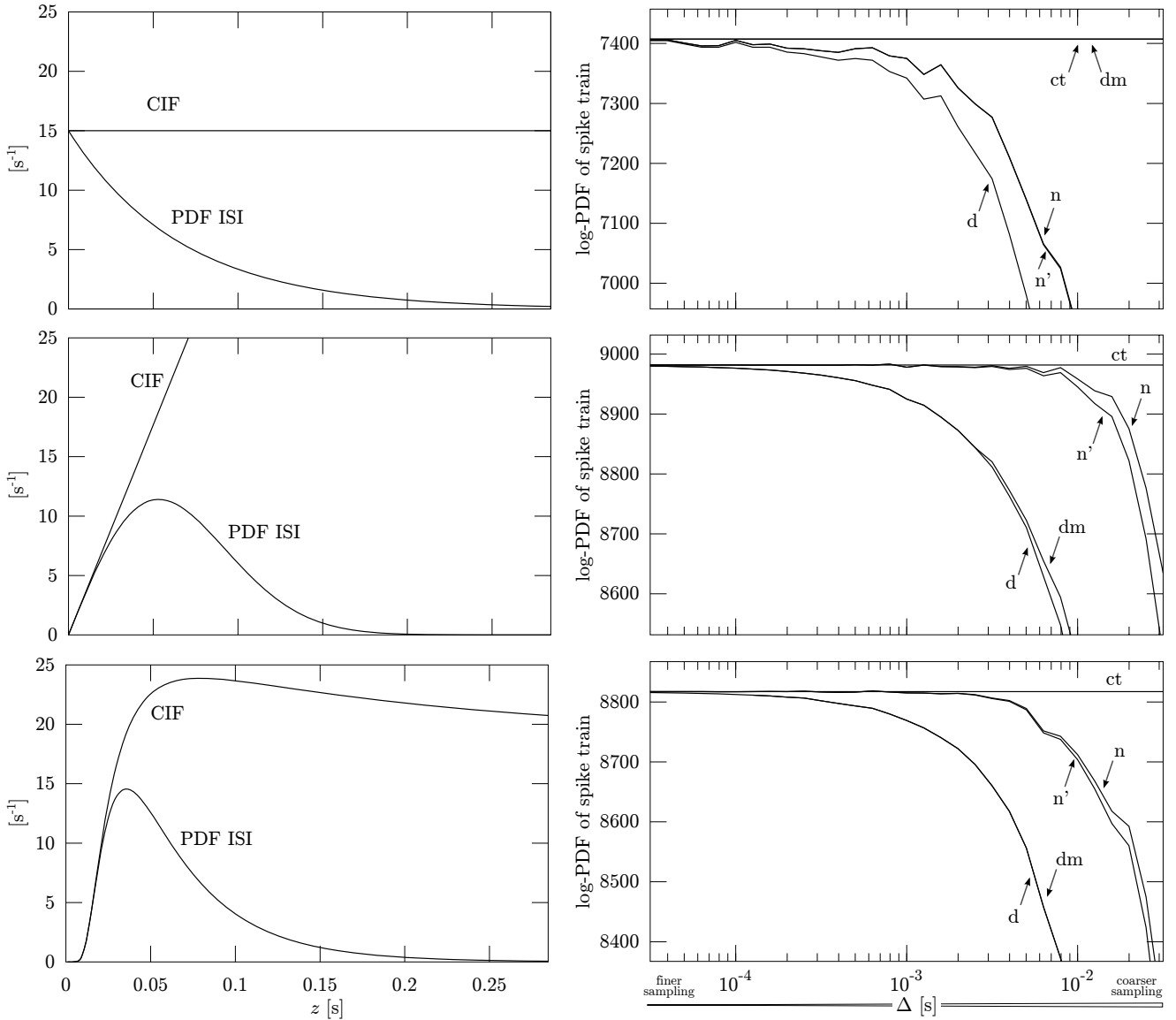


Figure 2: The plots on the left show the CIF of the point process and the PDF of the associated ISI distribution for the three examples: 1) homogeneous Poisson process with exponentially distributed ISIs (top); 2) Rayleigh-distributed ISIs (centre); 3) inverse-Gaussian-distributed ISIs (bottom). On the right, for each case, the plots show the corresponding log-PDF of the whole spike train evaluated from the simulated data (see Section 2.5) using: “ct”, the continuous time version (18); “n”, our discrete time definition (7); “n'”, its approximation (11); “d”, the conventional discrete time formulation (12); and “dm”, a variant of (12) allowing for more events per bin.

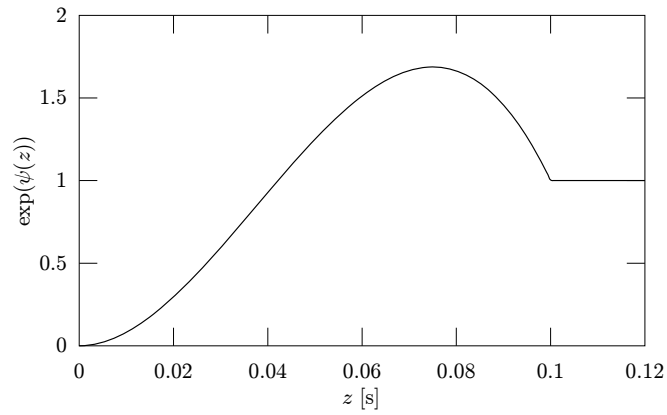


Figure 3: The plot represents the exp of the kernel function in (24), i.e., a continuous function in $(0, +\infty)$ that assumes value $c_1(z/c_3)^3 + c_2(z/c_3)^2$ if $0 < z < c_3$ and one otherwise.

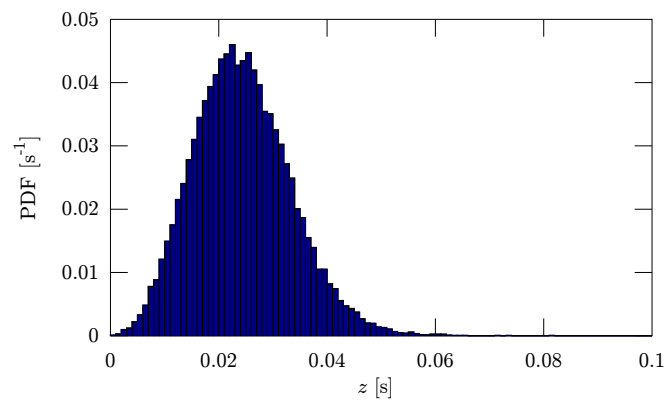


Figure 4: Empirical PDF (histogram) of the ISIs of the simulated neural spike train generated using the procedure described in Section 3.3.

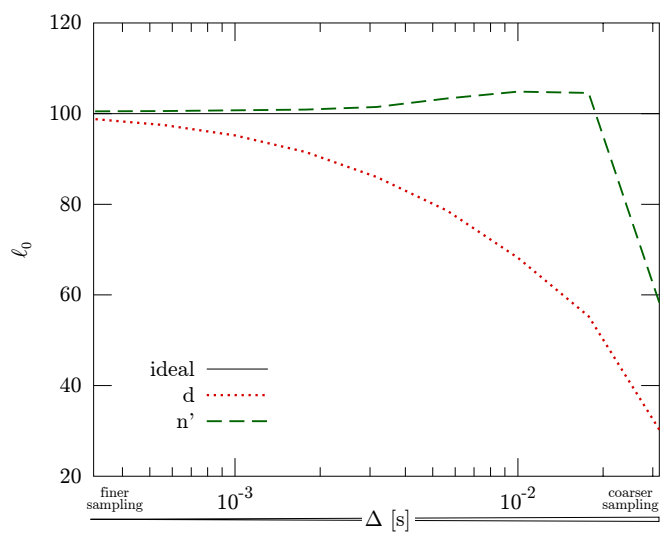


Figure 5: Value of ℓ_0 estimated using the conventional GLM (red dotted line) and using our log-linear model (green dashed line) as a function of the bin size, Δ .

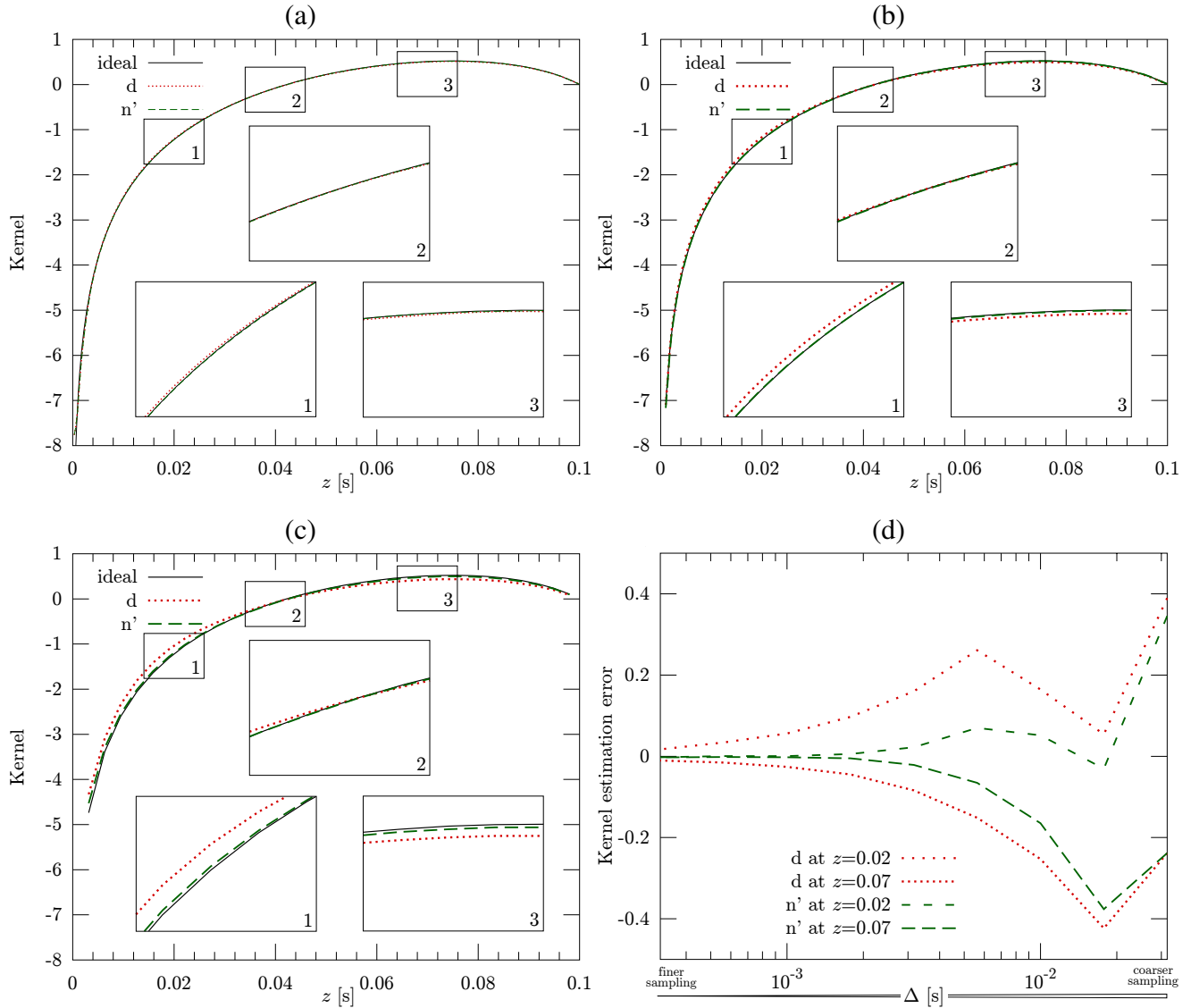


Figure 6: The figure compares the accuracies of the conventional GLM and of our log-linear model in the estimation of the autoregressive kernel $\psi(z)$ (see equation (24)). The subfigures (a)–(c) show the reconstructed kernel functions (represented by red dotted lines for the traditional GLM and green dashed lines for our log-linear model) and the target function $\psi(z)$ (black solid lines) for three values of bin sizes: $\Delta \simeq 0.3$ ms, $\Delta \simeq 1$ ms, and $\Delta \simeq 3$ ms, respectively. Subfigure (d) shows the estimation error (difference between the estimated value and the ideal one) for $z = 0.02$ s and $z = 0.07$ s (i.e., the centres of the zoom insets 1 and 3 in subfigures (a)–(c)) as a function of the bin size, Δ .

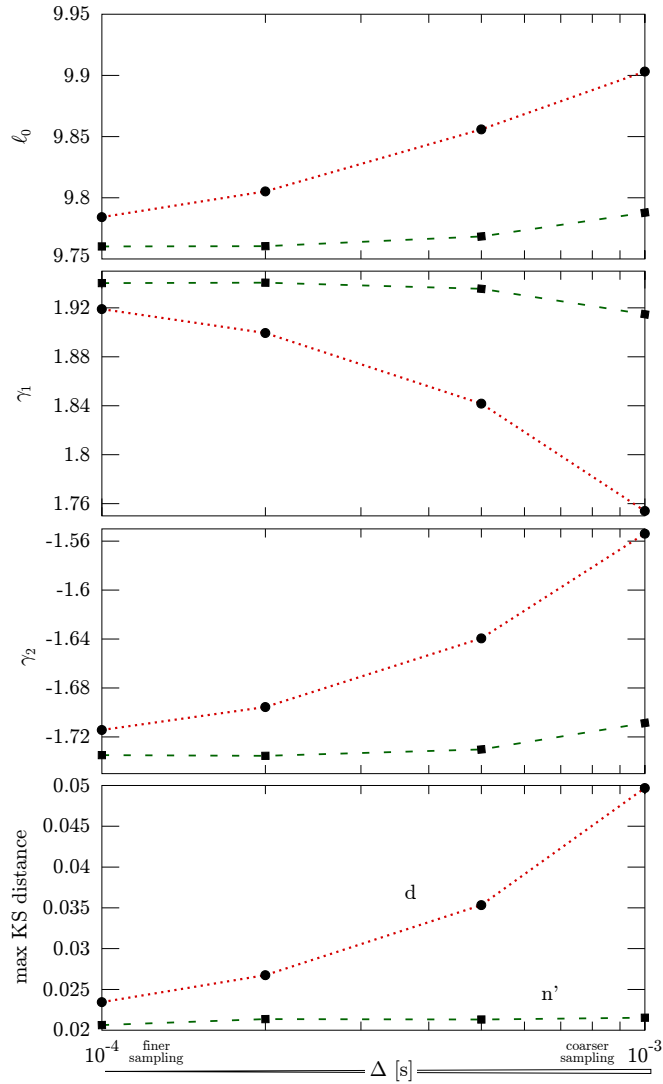


Figure 7: The top three subfigures show the results of the estimation of the parameters β_0 , γ_1 , and γ_2 of (27) as a function of the bin size when using conventional GLM (circles connected by red dotted segments) and our log-linear model (squares connected by green dashed segments). The bottom subfigure shows the KS-distance achieved by the fitted models for different bin sizes.