Non-parametric Approach to Extract Information from Interspike Intervals

Enrico Rossoni[†] Jianfeng Feng[‡]

[†]Department of Informatics, Sussex University, Brighton BN1 9QH, UK

[‡]Department of Computer Science, Warwick University

Coventry CV4 7AL, UK

http://www.cogs.susx.ac.uk/users/jianfeng

Corresponding to: Prof. Jianfeng Feng

February 25, 2005

Abstract

A nonparametric approach is developed to extract information from interspike interval data. In terms of Expectation-Maximization (EM) algorithm, interspike interval data from experiments and simulations are first approximated by a mixture of various probability distributions, including Gamma, inverse Gaussian, log-normal, and the interspike interval distribution of the leaky integrate-and-fire model. We demonstrate that our approach is successful when fitting benchmark data which failed to be fitted in the literature. Also, an approach to fit mixture distributions to censored data, collected naturally in trial-to-trial or multi-electrode array experiments, is presented. The software to perform above computations is available at http://www.informatics.sussex.ac.uk/users/er28/em/. Our results demonstrate how to efficiently and rapidly read out information from an ensemble of spike trains.

1 Introduction

The first and most important task in analyzing data recorded from neurons in nervous systems is possibly to fit the histogram of the interspike interval histogram (ISIH) with some known distribution densities[3, 5, 6, 7, 8, 17, 18, 24, 27]. In the literature, usually a single distribution (parametric approach) rather than a mixture of distributions (semi-parametric or non-parametric) are used[24, 29, 36]. As a consequence, many results reported in the literature fail to achieve a successful fitting, according to some statistical criteria such as the Kolmogorov-Smirnov (K-S) test.

This approach, termed parametric approach, in comparison with semi- and non-parametric approach in statistics, is possibly the most widely used one in dealing with experimental data, as amply demonstrated in the literature[1, 10, 13, 16, 21, 28, 33, 35]. In parametric approach, a distribution is given in priori and there are usually only a few parameters to be estimated, using some optimal criteria such as the maximum likelihood estimate. Of course, the parametric approach is over simplified and in most, if not all, practical applications, we have to use a semi- or non-parametric approach where we have to use a mixture of distributions.

A typical and well known example is the data recorded from neurons in the retina of the goldfish. In [24, 2], the authors fitted the data with an inverse Gaussian distribution and found that it achieved a better result than a Gamma distribution. However, the fitting fails to pass a statistical criterion such as the K-S test. The problem, realized by various authors [2], lies in the fact that the experimental histogram cannot be reduced to a single, easily to handle probability distribution. It is natural to expect that using a mixture of distributions will improve the fitting. However, the problem of parameter estimation for mixture models is not straightforward. Most of the time, explicit solutions for maximum likelihood estimates do not exist, and a numerical approach has to be adopted. Nevertheless, in Statistics in the past few years the EM algorithm [14] has been the most successfully used iterative approach to the estimation of mixture distribution parameters. Here, we employ the EM algorithm to fit interspike intervals data with a mixture of various

distributions, including Gamma, inverse Gaussian, log-normal and the interspike interval distribution of the leaky integrate-and-fire model. In fact, our approach could be easily adopted to other form of probability distributions as well.

We then test our approach for both experimental data and data generated from a biophysical model (Hodgkin-Huxley model) with stochastic inputs. For the Goldfish data, we found that our fitting (a mixture of one Gamma and two inverse Gaussian distributions) successfully passes the K-S test with a given p-value. As mentioned before, in the literature, no such a successful fitting is reported, although many people have tried to fit the data. For the Hodgkin-Huxley model, we first used a mixture of one Gamma and two inverse Gaussian distributions to successfully fit the data. Based upon our understanding of the mechanisms of generating the spikes of the Hodgkin-Huxley model, we then reduce the three distributions to a mixture of log-normal and the interspike interval distribution derived from the leaky integrate-and-fire model.

In physiological experiments, we usually repeat one experiment with identical setup by many trials. However, in each trial, either we only have a limit time window to record the data, or more realistically the recorded data is nonstationary and we have to analyze the data in a very short time window. This scenario also corresponds to the decoding problem. A down-stream neuron receives inputs from a group of neurons within a time bin, say 100 msec. Depending on the obtained information, the neuron has to decipher the meaning of the input. It is a practically and theoretically important issue to estimate the true parameters (input information) from truncated data. Here we borrow the idea from statistics of censored data and develop EM algorithms to tackle the problem. With simulated data, our approach successfully recovers the original information.

One might argue that to fit interspike intervals by probability distributions essentially lost important information in temporal domain. However, it is not true if we look at the problem from population coding point of view. As we mentioned above, if the time evolution of a population of spike trains are broken down to moving time windows. In each time window, we can fit the data via a few parameters including the weights of the mixtures. Then the whole information of the popu-

lation of the spike trains are summarized by the dynamics of the parameters and the time domain information can be fully recovered. In the current, we include two examples to confirm our claims. We can also apply our approach to data collected from multi-electrode array (MEA) experiments [23].

Software to fit interspike interval distributions with mixture of distributions is available at http://www.informatics.sussex.ac.uk/users/er28/em/.

2 Models

2.1 Probability distributions

First of all we list a few probability distributions which will be used in the current paper. Of course, we can easily generalize our approach to other distributions as well [11, 19, 25, 26].

1. The gamma probability density

$$p_{\mathrm{Gam}}(t; a, b) = \frac{b^a}{\Gamma(a)} t^{a-1} \exp[-t/b] \tag{1}$$

is a probability model frequently used to describe spike trains. Many data obtained from experiments can be roughly approximated by a single gamma distribution. The exponential probability density, which is the interspike interval model associated with a simple Poisson process, is a special case of Eq. (1) for $a = 1$. A Poisson process is memoryless (Markov). When $a \neq 1$, the spike trains have memory.

2. Another commonly used probability distribution is the inverse Gaussian density defined by

$$p_{\mathrm{IG}}(t; \mu, \lambda) = \left(\frac{\lambda}{2\pi t^3}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{\lambda(t-\mu)^2}{\mu^2 t}\right]$$

A renewal process model with an inverse Gaussian density as interspike interval distribution can be derived from a stochastic integrate-and-fire model in which the membrane voltage is represented as a random walk with drift [19], but *without* decay. This (over)-simplified model was first suggested by

Schroedinger [32] and was first applied in spike train data analysis by Gerstein and Mandelbrot [19].

3. We have recently derived the distribution of the interspike intervals of the leaky integrate and fire model [16]

$$p_{IF}(t|(\gamma,\mu,\sigma)) = \frac{2\sigma^2\gamma\mu\exp(-t/\gamma)}{\sqrt{\pi[\sigma^2\gamma(1-\exp(-2t/\gamma))]^3}} \cdot \exp\left[-\frac{(\gamma\mu)^2\exp(-2t/\gamma)}{\sigma^2\gamma(1-\exp(-2t/\gamma))}\right] \tag{2}$$

where γ is the membrane decay constant, μ and σ are mean input and input variance (see Subsection 3.2, Eq. (17)).

In the literature, the Gamma and inverse Gaussian distribution densities are most widely used to fit experimental data. One basic reason to favor them is that they have a very clear analytical expression. The other reason is because the gamma and inverse Gaussian ISI probability densities can be derived from elementary stochastic integrate-and-fire models, as mentioned above[2, 20, 15]. In fact, this claim is wrong and widely spread in the literature.d in the literature. The correct form is Eq. (2) which is much more complex than a Gamma or an inverse Gaussian distribution.

Therefore, the model we consider in the current paper will take the form

$$p(t|\Theta) = \sum_{k=1}^{K} w_k p_k(t|\theta_k), \tag{3}$$

where $p_k \in \{p_{Gam}, p_{IG}, p_{IF}, p_{LN}\}$, θ_k is the vector of parameters for each distribution and p_{LN} is for log-normal, K is the number of modules, $w_k > 0$, and $\sum_{k=1}^{K} w_k = 1$. The vector $\Theta = (w_1, \cdots, w_K, \theta_1, \cdots, \theta_K)$ represents the complete set of parameters for the mixture model. Hence the central topic of our current paper is to find the parameter Θ in terms of certain optimal criteria.

Given an observed set of ISIs, $\{t_1, t_2, \cdots, t_N\}$, the maximum likelihood estimate is applied to find the parameter Θ. the goal of the maximal likelihood estimation (MLE) is to estimate[30, 31, 34, 4] Θ that maximizes the likelihood function,

$$L(\Theta) = \prod_{n=1}^{N} p(t_n|\Theta)$$

or, equivalently, the log-likelihood function

$$\ell(\Theta) = \sum_{n=1}^{N} \log p(t_n|\Theta)$$

To find the maximal likelihood estimation above, we borrow the well known EM algorithm here, as summarized and re derived to our cases below[14].

2.2 EM for mixture distributions

It is easy to see that Eq. (3) is the marginal distribution of t of the joint distribution of $y = (t, z)$, i.e.

$$\text{Prob}(z = k|\Theta) = w_k, \quad k \in \{1, \ldots, K\},$$

and

$$t|(z, \Theta) \sim p_z(t|\theta_z)$$

Given a sample of size N, we treat $\{(t_i, z_i) : i = 1, \ldots, N\}$ as the complete data, where the value of z_i is not observable or missing variable. The EM algorithm is originally developed to estimate the missing variable. The log-likelihood of the complete-data is then,

$$\ell(\Theta) = \sum_{i=1}^{N} \log w_{z_i} + \sum_{i=1}^{N} \log p_{z_i}(t_i|\theta_{z_i})$$

Given the current estimate $\Theta^{(c)}$ and the observed values $\{t_i : i = 1, \ldots, N\}$, the expected complete-data log-likelihood is

$$Q(\Theta|\Theta^{(c)}) = \sum_{i=1}^{N} \left(\sum_{k=1}^{K} \hat{\alpha}_{i,k} \log w_k \right) + \sum_{i=1}^{N} \left(\sum_{k=1}^{K} \hat{\alpha}_{i,k} \log p_i(t_i|\theta_i) \right), \tag{4}$$

where

$$\hat{\alpha}_{i,k} = \frac{w_k^{(c)} p_k(t_i|\theta_k^{(c)})}{\sum_{j=1}^{K} w_j^{(c)} p_j(t_i|\theta_j^{(c)})} \tag{5}$$

for $k = 1, \ldots, K$ and $i = 1, \ldots, N$. The EM algorithm iteratively maximizes the function $Q(\Theta|\Theta^{(c)})$. More specifically, each iteration of EM consists of the following two steps:

- Compute $Q(\Theta|\Theta^{(c)})$ given the current estimate $\Theta^{(c)}$.

- Update Θ by maximizing $Q(\Theta|\Theta^{(c)})$ over Θ, that is

$$\hat{w}_k = \frac{1}{N}\sum_{i=1}^{N}\hat{\alpha}_{i,k}, \quad k = 1, \ldots, K, \tag{6}$$

and

$$\hat{\theta} = \arg\max_{\theta}\sum_{i=1}^{N}\sum_{k=1}^{K}\hat{\alpha}_{i,k}\log p_k(t_i|\theta_k), \tag{7}$$

where $\theta = (\theta_1, \ldots, \theta_K)$.

When $p_k(t|\theta_k)$ belongs to the exponential family, the equation (7) usually has closed-form solutions, e.g. for the Gaussian and the exponential distribution. For the Inverse Gaussian distribution, Eq.(7) has a particularly simple form, namely[1]

$$\begin{cases} \mu_j^{(c+1)} & = & \langle t \rangle_j^{(c)} \\ \\ \dfrac{1}{\lambda_j^{(c+1)}} + \dfrac{1}{\mu_j^{(c+1)}} & = & \left\langle \dfrac{1}{t} \right\rangle_j^{(c)} \end{cases} \tag{8}$$

where we defined, for a function F,

$$\langle F(t) \rangle_j^{(c)} = \frac{\sum_i \hat{\alpha}_{i,j}^{(c)} F(t_i)}{\sum_i \hat{\alpha}_{i,j}^{(c)}}$$

For the Gamma distribution we get

$$\begin{cases} a_j^{(c+1)} b_j^{(c+1)} & = & \langle t \rangle_j^{(c)} \\ \\ \Psi(a_j^{(c+1)}) - \log b_j^{(c+1)} & = & \langle \log t \rangle_j^{(c)} \end{cases} \tag{9}$$

where Ψ is the digamma function[36].

[1]To obtain this result, we recurred to a useful properties of the Inverse Gaussian distribution, namely that, for an IG sample, $(X_1, X_2, \ldots, X_n)$, the variates $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $V = \{\sum_{i=1}^{n}(1/X_i - 1/\bar{X})\}/n$ are independently distributed, and $n\lambda V$ is distributed as a χ^2_{n-1}, see [11].

2.3 Trial-to-trial Experiments

When spike data are collected in trial-to-trial experiments, measurements of interspike intervals are *censored* due to the finite duration of the recording period, $T_{\rm obs}$, i.e. only intervals with $t < T_{\rm obs}$ can be observed.

Depending on the probability distribution of the ISI, and on the severity of the censoring, it is possible that during many trials only one spike is observed, thus making impossible to define an interspike interval. Besides, even when two or more spikes are observed during a trial, the partial information about the last, censored, interval is usually discarded. On the other hand, it is evident that any statistical estimate which is based only on the *regular* intervals, will be inevitably subject to a bias, as we have demonstrated in [16]. In fact, the reason to introduce the bias is obvious. It is simply due to the limit of the observation time window: spikes longer than the given time window will never be collected.

The problem considered in [16] seems different from the issue we discussed here. Consider a group of neurons, they all send their spikes to a down stream neuron. We ask how can we decode the input information, for example, the input firing rate, in terms of the spikes received by the down stream neuron. We further require that the input information should be decoded as soon as possible, as in the nervous system. Hence the scenario is exactly we encounter here: we or the neuron has received spikes from a group of neurons within a time bin (T_{obs}) and we (the neuron) has to read out the information.

In order to correct this bias due to incomplete data, we resort to the theory of censored data in statistics. Let us first consider the case of a data set drawn from a single distribution, $p(t;\theta)$. The data set is partitioned into two classes: the *regular* intervals, $\{t_i\}_{i\in R}$, and the *truncated* intervals, $\{t_i\}_{i\in T}$ (in red in Fig. 4, upper panel), where the time of the end spike is replaced by $T_{\rm obs}$. This corresponds to the case of right-censoring discussed in [12]. Accordingly, the log-likelihood of the full data set is written as

$$L(\theta) = \sum_{i\in R} p(t_i;\theta) + \sum_{i\in T} S(t_i;\theta) \tag{10}$$

where S is the survival function,

$$S(t|\theta) = \int_t^{+\infty} p(t'|\theta)dt'.$$

For mixture of distributions, the log-likelihood of the complete data (z_i, t_i), in presence of censored observations, can be written as

$$L(\Theta) = \sum_{i\in R} \log w_{z_i} + \sum_{i\in R} \log p_{z_i}(t_i|\theta_{z_i}) + \sum_{i\in T} \log w_{z_i} + \sum_{i\in T} \log S_{z_i}(t_i|\theta_{z_i}) \quad (11)$$

To calculate the expectation of the expression above, we need to work out the conditional probability density of the hidden variable given the data $\{t_i\}_{i\in R} \cup \{t_i\}_{i\in T}$, and the current estimates for the parameters $\Theta^{(c)}$. For the regular intervals, these are again given by Eq.(5); for the truncated intervals, the probability of the ith observation having been generated by module k, cannot be obtained likewise, since we do not know the exact value of the interval but only a lower bound for it. However, we can use the Bayes theorem to get

$$\hat{\alpha}_{i,k} = \frac{w_k^{(c)} S_k(t_i|\theta_k^{(c)})}{\sum_{j=1}^K w_j^{(c)} S_j(t_i|\theta_j^{(c)})} \quad \text{when} \quad i \in T \quad (12)$$

Therefore, in presence of censored observations, Eq.(4) must be rewritten as

$$\begin{aligned} Q(\Theta|\Theta^{(c)}) = & \sum_{i\in R}\sum_{k=1}^K \hat{\alpha}_{i,k} \log w_k + \sum_{i\in R}\sum_{k=1}^K \hat{\alpha}_{i,k} \log p_i(t_i|\theta_i) + \\ & \sum_{i\in T}\sum_{k=1}^K \hat{\alpha}_{i,k} \log w_k + \sum_{i\in T}\sum_{k=1}^K \hat{\alpha}_{i,k} \log S_i(t_i|\theta_i), \end{aligned} \quad (13)$$

By maximizing the expression above with respect to Θ, we obtain the recurrence formula for the parameters,

$$\hat{w}_k = \frac{1}{N}\left(\sum_{i\in R} \hat{\alpha}_{i,k} + \sum_{i\in T} \hat{\alpha}_{i,k}\right) \quad k = 1, \ldots, K, \quad (14)$$

$$\hat{\theta} = \arg\,\max_\theta \left(\sum_{i\in R}\sum_{k=1}^K \hat{\alpha}_{i,k} \log p_i(t_i|\theta_i) + \sum_{i\in T}\sum_{k=1}^K \hat{\beta}_{i,k} \log S_i(t_i|\theta_i)\right) \quad (15)$$

Despite the fact a closed form for the solution of (15) does not exist in the general case, the problem can be tackled numerically with relative ease, given the independence of each module's parameters.

3 Results

3.1 Experimental Data

As a first example, we considered a data set composed of $N = 975$ interspike intervals recorded from a goldfish retinal ganglion cell *in vitro*. Recordings were made with an extracellular microelectrode under constant illumination. The spike activity of this neuron reveals a collection of short and long ISIs, as reflected in the apparently bimodal character of the ISIH. To fit these data we consider mixtures of 2, and 3 modules. Goodness of fit has been assessed by a two-sided two-sample K-S test, where the experimental data set is compared with a simulated data set generated from the estimated model. When mixtures of two modules were considered, all estimated models failed to pass the K-S test ($P < 0.05$), regardless of the distribution types considered. The goodness of fit is improved when considering mixtures of 3 modules, with some of the models meeting the criterion for acceptance, see Tab. 3.1. In Fig.1 (left) we compared the ISIH from the experimental data with the expected frequencies calculated from the estimated model with the highest P-value. A plot of the respective empirical cumulative distribution functions is shown in Fig.1 (right). It could be easily seen that our approach fits the data well. To the best of our knowledge, it is the first time in the literature that a fitting of the data passes the KS test.

The fitted parameters are $(a, b) = (15.1673, 0.3243)$ for the Gamma distribution, $(\mu_1, \lambda_1) = (13.6612, 24.9184)$ and $(\mu_2, \lambda_2) = (90.9478, 750.7258)$ for the two inverse Gaussian distribution, with $(w_1, w_2, w_3) = (0.2592, 0.4912, 0.2497)$. More specifically we have (see Fig. 1)

$$\begin{aligned}\hat{p}(t) &= \frac{0.2592 \times 0.3247^{15.1673}}{\Gamma(15.1673)} t^{14.1673} \exp[-t/0.3243] \\ &\quad +0.4912 \left(\frac{24.9184}{2\pi t^3}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{24.9184(t-13.6612)^2}{13.6612^2 t}\right] \\ &\quad +0.2497 \left(\frac{750.7258}{2\pi t^3}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{750.7258(t-90.9487)^2}{90.9487^2 t}\right]\end{aligned} \tag{16}$$

To further demonstrate the power of our approach, we turn our attentions to the classical Hodgkin-Huxley model.

Model	P	K-S
(1,1,-)	4.8e-4	0.098
(1,2,-)	6.1e-3	0.082
(2,1,-)	1.5e-3	0.091
(2,2,-)	2.6e-3	0.088
(1,1,1)	1.2e-3	0.092
(1,1,2)	0.033	0.069
(1,2,1)	0.072	0.062
(1,2,2)	0.063	0.063
(2,1,1)	**0.082**	**0.061**
(2,1,2)	0.028	0.070
(2,2,1)	0.013	0.076
(2,2,2)	0.028	0.070

Table 1: Goodness of fit of the estimated mixture models obtained by the EM algorithm. The calculated P values and the values of the K-S statistics for varying number and type of distributions, are shown for the case of experimental and simulated spike data (see Fig. 1). The best fit is reported in bold face type. 2=Gamma, 1=inverse Gaussian

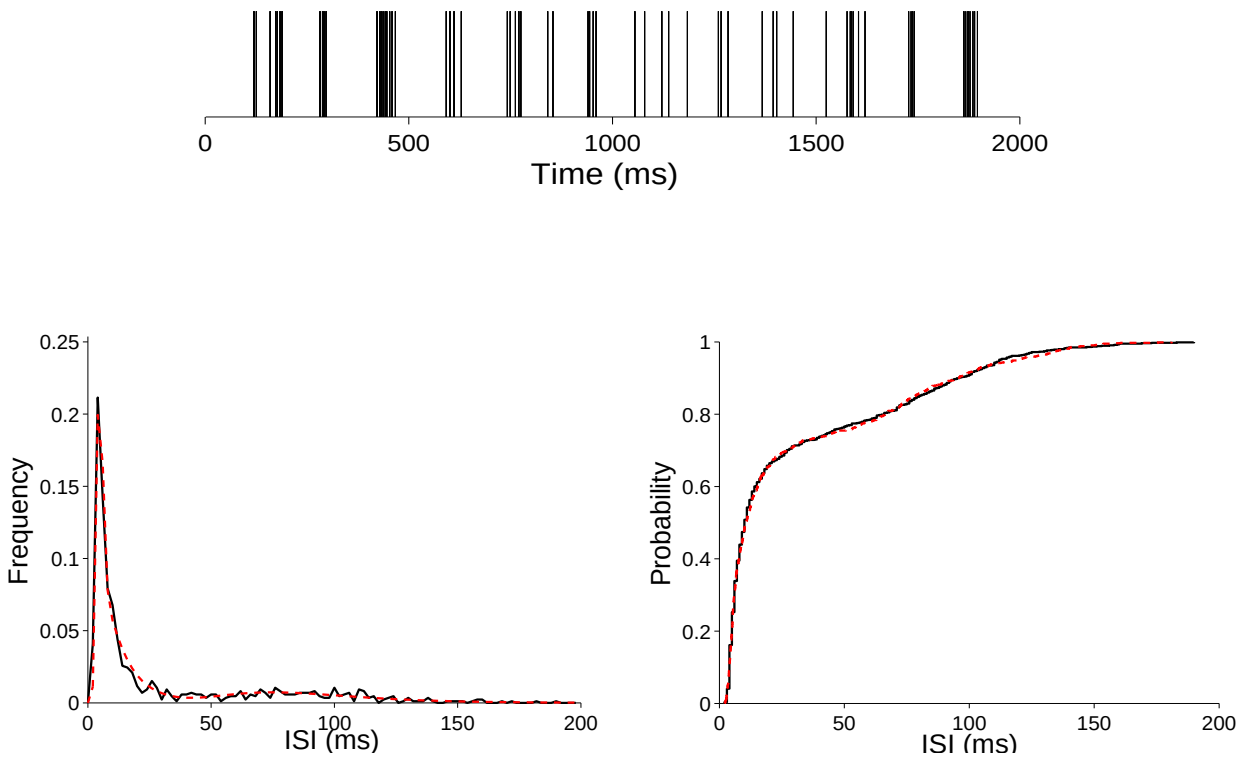

Figure 1: Fitting results of the gold fish data. **Upper panel** Spike trains from the gold fish. (**Bottom Left**) A comparison of the ISIH calculated for the experimental data set (solid line) and for the estimated mixture model with the highest P-value (dashed line, i.e. $\hat{p}(t)$ defined by Eq. 16). (**Bottom Right**) Comparison of empirical cumulative distribution functions.

3.2 HH model

We consider a sample of $N = 10000$ interspike intervals generated by simulating the response of the Hodgkin-Huxley model to a Gaussian input[2],

$$CdV = -g_{Na}m^3h(V - V_{Na})dt - g_k n^4(V - V_k)dt - g_L(V - V_L)dt + \mu dt + \sigma dB_t \quad (17)$$

$$\frac{dn}{dt} = \frac{n_\infty(V) - n}{\tau_n(V)}, \qquad \frac{dm}{dt} = \frac{m_\infty(V) - m}{\tau_m(V)}, \qquad \frac{dh}{dt} = \frac{h_\infty(V) - h}{\tau_h(V)} \quad (18)$$

where B_t is the standard Brownian motion, and we chose $\mu = 7$, $\sigma = 2$. For the parameters considered, the HH model has a fixed point, which corresponds to a depolarized stable resting potential, and a limit cycle, which is manifest as repetitive firing. As a result of the stochastic input, the solutions of the system (17)-(18) alternate randomly periods of repetitive firing, at about the limit cycle's unperturbed frequency, and underthreshold oscillations around the resting state, thereby originating a multimodal distribution for the interspike intervals (see Fig. 2). Again, we apply the EM algorithm to determine which mixture model best approximated our data. The performance of the estimated models are reported in Tab. 3.1; compared ISIH for the real data and the best-model as shown in Fig.2, together with the respective empirical cumulative distribution functions (Eq. (19)).

$$\begin{aligned} \hat{p}(t) \;=\; & \frac{0.3761 \times 0.1^{126.6}}{\Gamma(126.673)} t^{126.6} \exp[-t/0.1] \\ & +0.4091 \left(\frac{15.1}{2\pi t^3}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{15.1(t-4477.1)^2}{4477.1^2 t}\right] \\ & +0.2148 \left(\frac{28.8}{2\pi t^3}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{28.8(t-247.9)^2}{247.9^2 t}\right] \end{aligned} \quad (19)$$

Now let us have a more close look of the implication of our results. As we mentioned before, we fit our data by a mixture of distributions. It is certainly true that we can fit any set of data by increasing the number of mixtures. Of course, increasing the number of mixtures increases the complexity, as could be measured by, for example, AIC etc. Hence in our approach, we started from two, three, ... mixture of distributions and stopped whenever the statistical criterion is achieved, here is the K-S test. After successfully fitting the HH model with three

[2]We refer the reader to [22] for a detailed definition of the model and the values of the parameters

Model	P	K-S
(1,1,-)	1.4e-5	0.034
(1,2,-)	3.4e-8	0.042
(2,1,-)	5.7e-8	0.042
(2,2,-)	6.4e-6	0.035
(1,1,1)	0.11	0.017
(1,1,2)	0.0090	0.023
(1,2,1)	0.11	0.017
(1,2,2)	0.13	0.017
(2,1,1)	**0.20**	**0.015**
(2,1,2)	0.092	0.017
(2,2,1)	0.051	0.019
(2,2,2)	0.042	0.020

Table 2: Goodness of fit of the estimated mixture models obtained by the EM algorithm. The calculated P values and the values of the K-S statistics for varying number and type of distributions, are shown for the case of experimental and simulated spike data. The best fit is reported in bold face type.

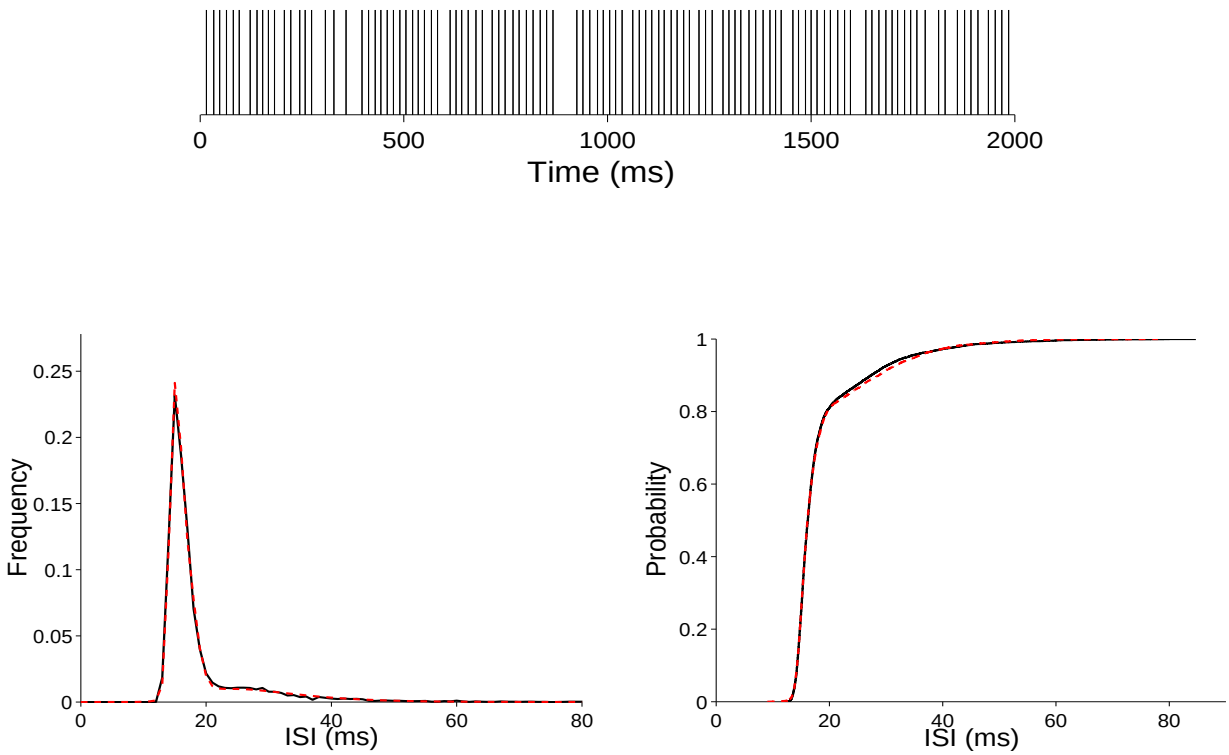

Figure 2: Fitting results for HH model. **Upper panel** Spikes train of the HH model. (**Bottom Left**) A comparison of the ISIHs obtained from a solution of system (17)-(18) (solid), and the estimated mixture model with the highest P-value (dashed). (**Bottom Right**) Comparison of empirical cumulative distribution functions.

distributions, we will naturally ask whether we could further reduce the number of distributions by varying the distribution formality. We know that in the parameter regions of input parameters to the HH model, there are two attractors. One is the limit cycles (spikes) and the other is a stable fixed point, corresponding to the resting potential. For limit cycles, with the presence of noise, we would expect that its interspike intervals is perturbed and should be distributed according to Gaussian type distribution. For the jumping between the fixed point and the limit cycle, from the large deviation theory we know that is should be exponentially distributed, or at least in the limit case. Therefore, we apply the mixture of a log-normal distribution and the distribution of the leaky integrate-and-fire model to fit the data. Fig. 3 clearly shows that using two distributions rather than three we can fit the data well.

3.3 Censored Data

Finally, we address the cases of fitting ISIH with censored observations: one corresponds trail-to-trial experiments and the other is MEA data.

3.3.1 Trial-to-trial Experiment

To this aim we consider a simulated data set composed of $N = 3000$ ISIs drawn from the distribution

$$p(t|\Theta) = \sum_{i=1}^{3} w_i p_{\mathrm{Gam}}(t|a_i, b_i), \tag{20}$$

with parameters $(a_1, a_2, a_3) = (5.21, 54.40, 25.43)$, $(b_1, b_2, b_3) = (84.04, 2.92, 5.37)$, and weights $(w_1, w_2, w_3) = (0.34, 0.11, 0.55)$. The obtained spike train has been split into $n = 1438$ chunks of length $T_{\mathrm{obs}} = 500$ ms each, to provide a trial-to-trial data set. From these data, we extracted $N_R = 1676$ regular and $N_T = 1324$ truncated intervals. Notice that, due to the severe censoring, more than 25% of the trials ($n = 362$) contained only one spike, as shown in the raster plot of Fig.4.

As shown in Fig. 4, bottom left, the standard EM algorithm provides a good approximation of the distribution of the regular ISI, but fails to account for the statistical features of the complete data set, i.e. the model Eq. (20). By introducing the censoring correction in the likelihood, the match between the ISIH of the

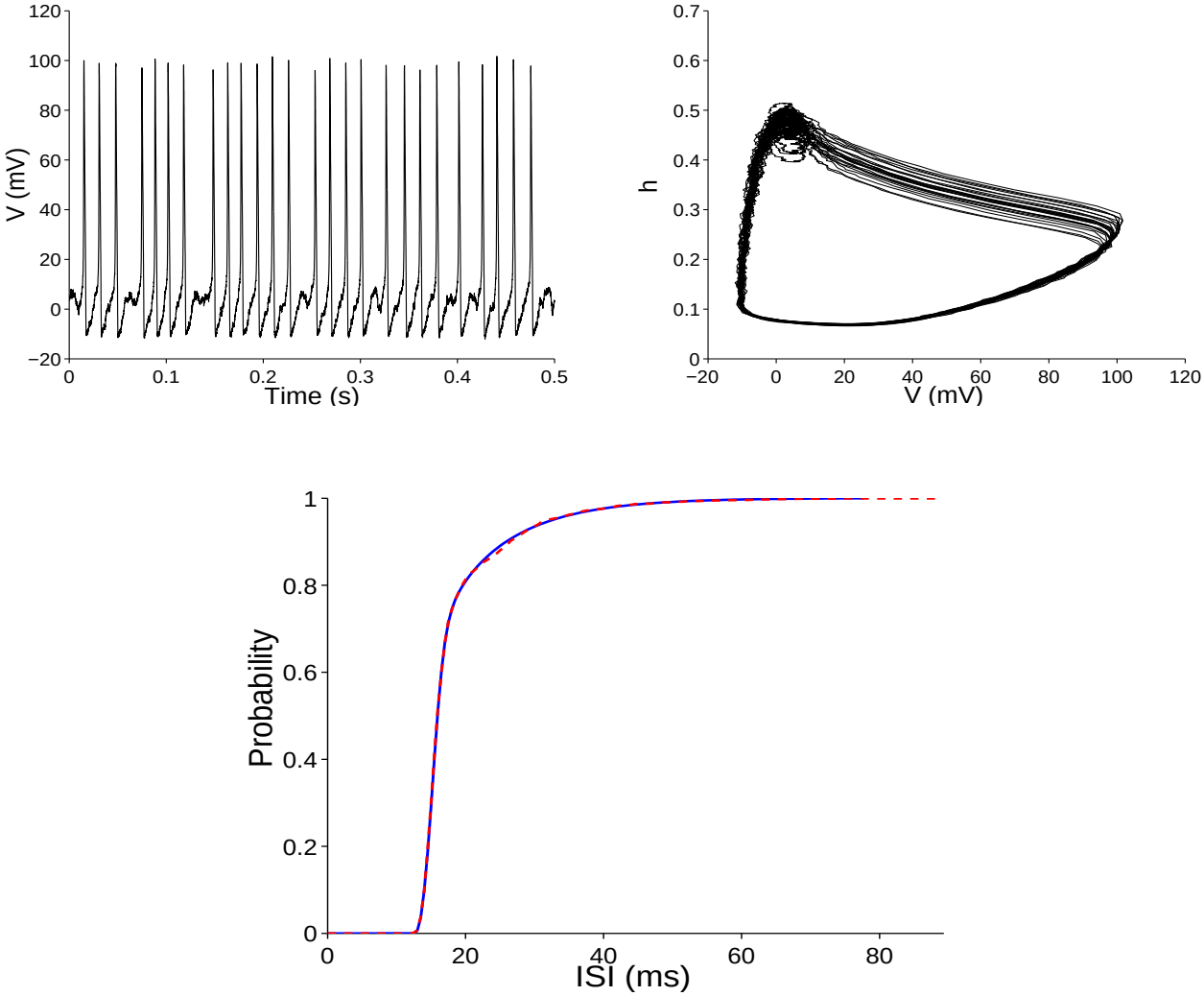

Figure 3: **Upper panel, left**, the trajectory of the membrane potential of the HH model; **right**, the trajectory plotted on the $h - V$ space. It is readily seen that the random perturbations cause the model exhibit a bursting activity. **Bottom**: fitting with a mixture of two distributions achieves satisfactory results (red line).

complete data set and of the reconstructed model is considerably improved, Fig.4, Right.

The parameters in fitted probability distribution are: without censoring correction,

$$\begin{aligned}(w_1, w_2, w_3) &= (0.62, 0.22, 0.16),\\ (a_1, a_2, a_3) &= (32.94, 85.13, 8.88),\\ (b_1, b_2, b_3) &= (3.90, 1.87, 29.42)\end{aligned}$$

and with censoring correction,

$$\begin{aligned}(w_1, w_2, w_3) &= (0.35, 0.30, 0.35),\\ (a_1, a_2, a_3) &= (37.61, 58.46, 4.97),\\ (b_1, b_2, b_3) &= (3.32, 2.68, 83.06).\end{aligned}$$

It is easily seen that although with censoring correction, we can approximate the original distribution well, the actual parameter values are different. This could be due to the local minimums in finding the optimal parameters.

3.3.2 Dynamical Signals

To further test our approach, we now turn our attention to dynamical signals. As we mentioned before, this naturally corresponds to the data recorded from MEA experiments[23].

The setup is schematically plotted in Fig. 5. A dynamical signal as the rate of independent Poisson processes is sent to a group (100) of the leaky integrate-and-fire model and the output spikes are recorded in 2S. The decoding windows are 50 msec and disjoint. It is easily seen that with censoring correction MLE, the original signal is correctly retrieved, but a usual MLE fails (see Fig. 5).

If we employ further constrains on the retrieved signals, for example, smoothness of the curve as in Bayesian approach, we can certainly improve the decoding quality. We will apply our approach to MEA data [23].

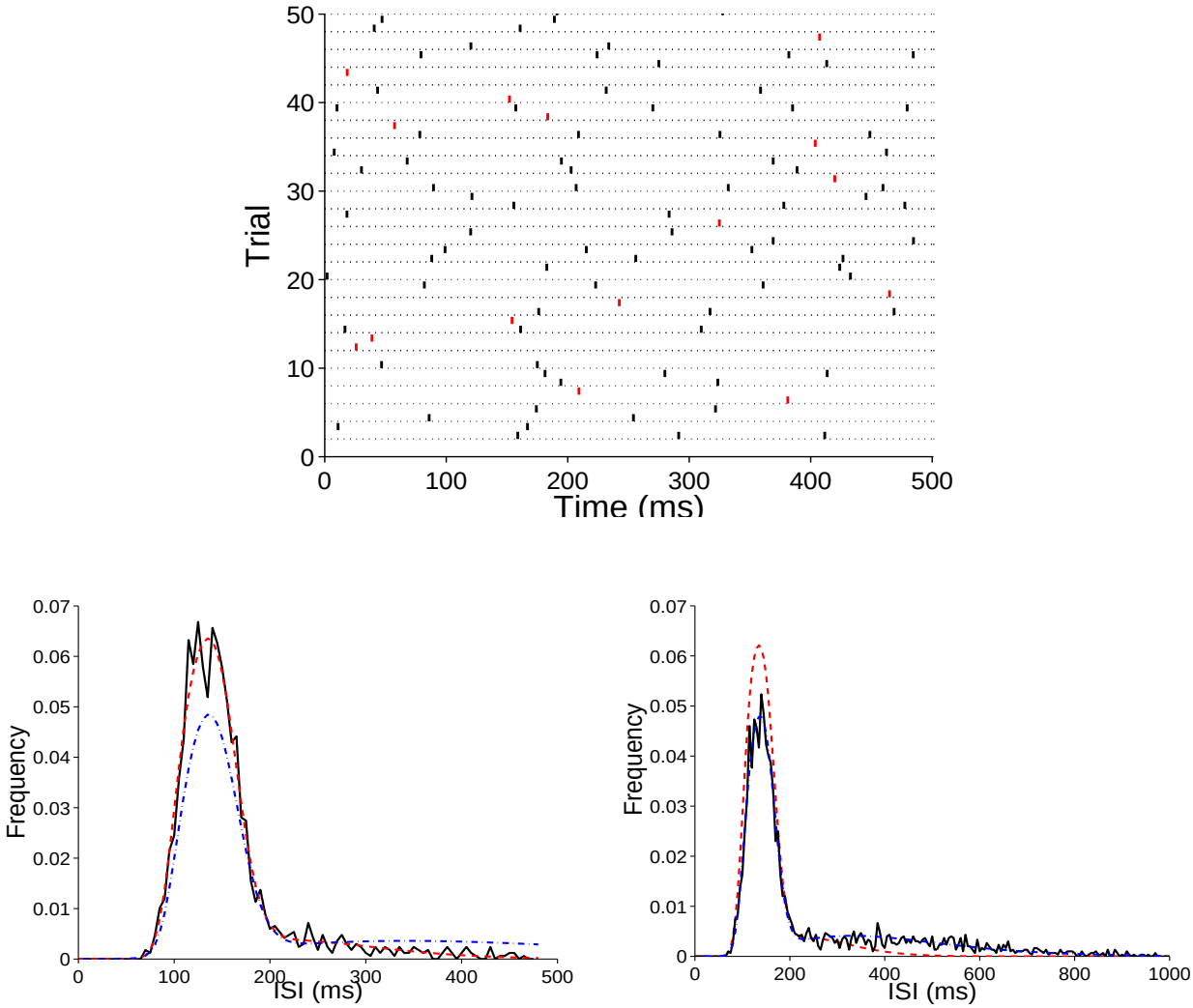

Figure 4: Upper panel, spike raster plots for the first 50 trials in the data set used for testing the censoring EM. Trials containing a single observation (not a complete interspike interval) have spikes drawn in red; horizontal dashed lines have been added for visual clarity. Bottom panel, left, The histogram of the *regular* interspike intervals generated by model (solid), compared with the frequencies calculated from the model estimated by the standard (dashed) and the modified EM algorithm (dot-dashed). Bottom panel, right, The histogram of the complete set of interspike intervals generated by model (solid), compared with the frequencies calculated from the model estimated by the standard (dashed) and the modified EM algorithm (dot-dashed).

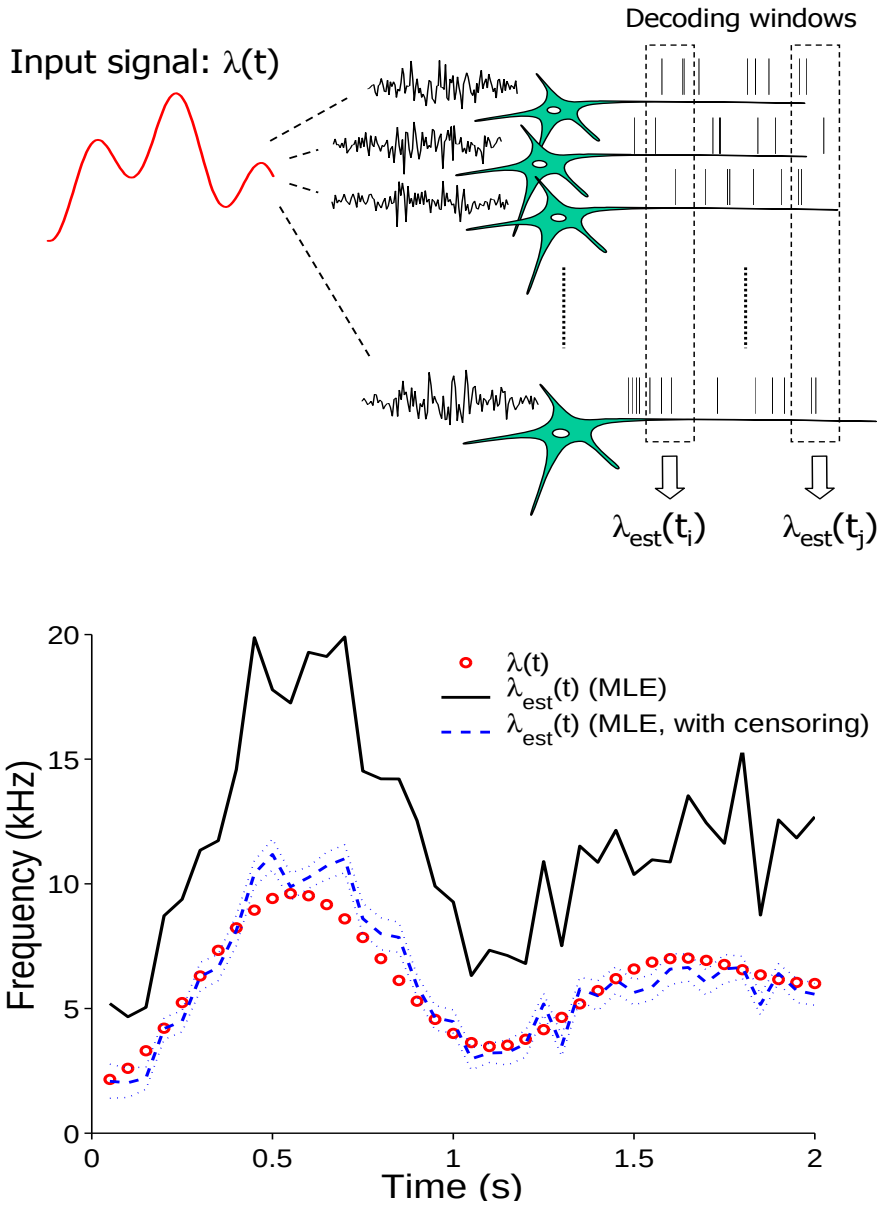

Figure 5: Decoding of dynamical signals from a group of neuron activity. Upper panel: a schematic plot of the setup. Bottom panel: results of one example with a population of 100 neurons. The input signal is $\lambda(t) = \lambda_0 + 2\lambda_0[\sin^2(\nu t) + \sin^2(0.75 \cdot \nu t)]$, with $\nu = \pi$ s^{-1}, $\lambda_0 = 2$ kHz. Parameters for the integrate-and-fire model are: $\gamma = 20$ms (membrane time constant),$a = 0.5$mV (PSP amplitude), $V_{th} = 20$mV (threshold). The decoding is done over disjoint stretches of 50ms; the dotted lines are 95% confidence intervals.

4 Discussions

We have presented a nonparametric approach to fit interspike interval distributions, with data from both experiments and simulations. Both censored and noncensored cases are considered.

4.1 Why fitting?

It would always not be satisfactory to approximate or to fit a quantity (here it is the distribution) rather than to accurately describe it. To fit a data is always the final way we will resort to. However, from our past few years experiences, we know that it is in general difficult, if it is not impossible, to find the exact distribution of interspike intervals, even for the simplest neuronal model, the leaky integrate and fire model, which is linear before resetting. When we face a biophysical model such as the HH model, we know it is totally impossible to derive an analytical formula for the interspike interval distribution. Certainly, we do not expect to have an analytical expression for the interspike interval distribution for a pyramidal cell in the cortex *in vivo*. Hence to use various approximations or fitting methods to explore the properties of interspike intervals is an efficient way to understand the neuronal dynamics.

Once we have a reasonable probability distribution to fit the data, we are then in an excellent position to investigate the neuronal mechanisms to generate spike and to extract the underling information, with the help of modern probability theory.

4.2 Reduction from high dimension to low dimension

When we record a hundred neurons over a time period T, we obtain a spatial temporal pattern of spikes. The pattern of spikes is of high dimension and it is usually difficult to grasp the essence of the pattern. However, after fitting a mixture of distributions to the data, we can greatly reduce the dimensionality of the data and we are thus in a much better position to extract information form the data. A typical example is shown in Subsection 3.3.2. Of course, in general, we would

expect the reduced dimension is more than one dimension, including weights and parameters in each distributions.

4.3 Decoding

One of the most challenging tasks in neuroscience is to understand how information is encoded and then decoded in the ensemble of spike trains. We have demonstrated how to apply our approach to such problems. The general results are very encouraging and we will adopt our approach to deal with experimental data, for example the MEA data from the olfactory bulb[23].

4.4 Temporal information

In the current paper, we treat each decoding window as independent (see Fig. 5). Of course we can easily include temporal relationship, as we mentioned above. Another way to include temporal relationship is to introduce nonhomogeneous point process as developed in [2]. It would be interesting to see how the current results can be improved if we combine two approaches together: to perform decoding with a nonhomogeneous point process and with a mixture of distributions as its interspike interval distribution.

4.5 Other approaches

In the literature, we have seen that there are a few other approaches to analyze and model the similar data we have here. We have tested some of them, for example, the Hidden Markov chain approach[9]. The obtained results are quite disappointing. Neuronal data, for example, those generated from the HH model, can not be well fitted by a Hidden Markov chain model. The hidden Markov chain model is tested in data collected from oxytocin and vasopressin neurons and again the outcome is not encouraging.

4.6 Uniqueness of parameters

Finally, as we pointed out above, the estimated parameters are usually not unique. This poses a serious problem for the physiological explanation of the fitted parameters. Technically, the problem is common for a complex problem such as we deal here: there are usually more than one solution.

Acknowledgment. We would like to thank E.N. Brown for his discussions on censoring data and providing us the Goldfish data. J.F. was partially supported by grants from UK EPSRC(GR/R54569), (GR/S20574), and (GR/S30443).

References

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

[2] Brown E.N., Barbieri R., Eden U.T., and Frank L.M. (2004) Likelihood methods for neural spike train data analysis, 253-286, in *Computational Neuroscience: A Comprehensive Approach* Feng J.F. (ed.), Chapman and Hall/CRC, Boca Raton.

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

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

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

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

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

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

[9] Camproux, A. C., Saunier, F., Chouvet, G., Thalabard, J. C., Thomas, G. (1996). A hidden Markov model approach to neuron firing patterns. *Biophys. J.*, **71**, 2404-2412.

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

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

[12] Cox P.R., and Lewis P.A.W. (1966) *The statistical analysis of series of events*, Latimer Trend & Co.Ltd. Whitstable.

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

[14] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser. B*, **39**, 1-38.

[15] Feng J.F. (2004) *Computational Neuroscience: A Comprehensive Approach*, Feng J.F. (ed.), Chapman and Hall/CRC: Boca Raton.

[16] Feng, J., and Deng, M. (2004). Decoding spikes in a spiking neuronal network. *J. Phys. A: Math. Gen.*, **37**, 5713-5727.

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

[18] Frank, L. M., Eden U. T., Solo, V., Wilson, M. A., and Brown, E. N., (2002). Contrasting patterns of receptive field plasticity in the hippocampus and the

entorhinal cortex: an adaptive filtering approach. *Journal of Neuroscience*, **22**, 3817-3830.

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

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

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

[22] Hodgkin, A. and Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500 - 544.

[23] Holton P., Nicol A., Kendrick K., and Feng J.F. (2005) MEANOVA–Application of MANOVA to MEA data, *J. Neurosci. Methods* (in press).

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

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

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

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

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

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

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

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

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

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

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

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

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