ELSEVIER

Journal of Neuroscience Methods 150 (2006) 30–40

JOURNAL OF NEUROSCIENCE METHODS

www.elsevier.com/locate/jneumeth

A nonparametric approach to extract information from interspike interval data

Enrico Rossoni [a], Jianfeng Feng [b,c,*]

[a] *Department of Informatics, Sussex University, Brighton BN1 9QH, UK*
[b] *Department of Mathematics, Human Normal University, 410081 Changsha, PR China*
[c] *Department of Computer Science, Warwick University, Coventry CV4 7AL, UK*

Received 10 March 2005; received in revised form 10 May 2005; accepted 23 May 2005

Abstract

In this work we develop an approach to extracting information from neural spike trains. Using the expectation-maximization (EM) algorithm, interspike interval data from experiments and simulations are fitted by mixtures of distributions, including Gamma, inverse Gaussian, log-normal, and the distribution of the interspike intervals of the leaky integrate-and-fire model. In terms of the Kolmogorov–Smirnov test for goodness-of-fit, our approach is proved successful ($P > 0.05$) in fitting benchmark data for which a classical parametric approach has been shown to fail before. In addition, we present a novel method to fit mixture models to censored data, and discuss two examples of the application of such a method, which correspond to the case of multiple-trial and multielectrode array data. A MATLAB implementation of the algorithm is available for download from http://www.informatics.sussex.ac.uk/users/er28/em/.

Keywords: Expectation-maximization (EM) algorithm; Multielectrode; Algorithm

1. Introduction

The first task in analyzing data recorded from the nervous system is to fit the histogram of the interspike intervals (ISIH) by some known probability densities (Brillinger, 1988; Iyengar and Liao, 1997; Brown et al., 1998; Frank et al., 2000; Barbieri et al., 2001; Kass and Ventura, 2001; Brown et al., 2001; Brown et al., 2002; Frank et al., 2002).

In a *parametric* approach to density estimation, it is assumed that the probability density can be represented as a specific functional form containing several adjustable parameters. Once such a functional form has been chosen, the parameters are estimated using optimal criteria, e.g. maximum-likelihood.

This approach is widespread in neuroscience data analysis, as amply demonstrated in the literature (Casella and Berger, 1990; Snyder and Miller, 1991; Levine, 1991; Andersen et al., 1993; Taylor and Karlin, 1994; Guttorp, 1995; Feng and Ding, 2004). Nevertheless, many results reported in literature failed to meet statistical criteria for goodness-of-fit such as the Kolmogorov–Smirnov (KS) test.

The problem, as realized by various authors, is that the interspike interval data cannot always be described in terms of a single probability density, as clearly illustrated, for instance, by the data recorded from the ganglion cells in the goldfish retina (Brown et al., 2004). In cases like this, it is reasonable to take a (semi-parametric) approach, in which the data are fitted by *mixtures* of distributions (McLachlan and Peel, 2000).

In the present work, we consider mixtures of distributions that include Gamma, inverse Gaussian, log-normal and the distribution of the interspike intervals generated by the leaky integrate-and-fire model. The problem of parameter estimation for such mixture models is not straightforward, since explicit formulas for maximum-likelihood estimates do not exist. However, a solution to this problem can be derived in terms of the expectation-maximization (EM) algorithm (Dempster and Laird, 1977), as we discuss in Section 2.2. A software implementation of this algorithm is tested on

* Corresponding author. Tel.: +44 7910660548.
E-mail address: Fengjf@hunnu.edu.cn (J. Feng).
URL: http://www.dcs.warwick.ac.uk/∼feng.

0165-0270/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2005.05.021

benchmark data sets from experiments and simulations. In particular, we find that the goldfish ganglion cell data can be successfully fitted by a three-component mixture model (KS test, $P > 0.05$). Also, we find that a two-component mixture can be fitted accurately to the distribution of the ISI generated by the Hodgkin–Huxley model with a stochastic input (KS test, $P > 0.05$).

An additional issue addressed in the current paper, is how to fit interspike interval data that contain *censored* observations. Although this receives only limited attention in the literature, the problem is of paramount importance for fitting data collected in multiple-trial recordings, especially when the spike activity can be recorded for only short periods of time, due to technical limitations or to nonstationarities of the data. The same problem arises in the context of neural population decoding. Indeed, in order to reconstruct the time course of a rapidly changing stimulus from a spike train ensemble, one is usually forced to consider short time windows during which the stimulus is approximately constant. Under these circumstances, a potentially large proportion of interspike intervals may not be observable, since they exceed the trial length. Besides, in cases when typical interspike intervals are longer than the period of observations, the majority of the trials may contain less than two spikes, thus only very limited information is available about the distribution of the interspike intervals. However, even in these cases it is still possible to obtain accurate estimates of the underlying ISI probability density if one takes into account also the truncated (censored) intervals, which are usually discarded. This approach is discussed in Section 2.3. In particular, we show how to modify the EM algorithm to estimate the parameters of a mixture model from data sets that include both regular and truncated interspike intervals. Finally, an example of application of this approach to a simple decoding task is illustrated in Section 3.3.2.

2. Methods

2.1. Mixture models

A K-component mixture model is defined by the probability density

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

where p_k are the component densities, with θ_k the vector of parameters for each component, and w_k are the mixing proportions, or weights, with $w_k > 0$ and $\sum_{k=1}^{K} w_k = 1$. The vector $\Theta = (w_1, \ldots, w_k, \theta_1, \ldots, \theta_k)$ represents the complete set of parameters for the mixture model.

Below, we review the probability densities that will be used in the current paper.

(1) The gamma probability density

$$p_{\mathrm{Gam}}(t; a, b) = \frac{b^{-a}}{\Gamma(a)} t^{a-1} \exp\left(\frac{-t}{b}\right) \tag{2}$$

is frequently used to model spike data, as many experimental ISIH have been shown to be roughly approximated by a single gamma distribution. It can be shown that (2) is the probability density of the ISI obtained from a simple stochastic integrate-and-fire model where the inputs are Poisson processes with constant rate (Tuckwell, 1988). Note that the exponential probability density, which is the distribution of the interspike intervals associated with a simple Poisson process, is a special case of Eq. (2) for $a = 1$.

(2) The inverse Gaussian probability density with parameters λ and μ

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

is the probability density of the interspike intervals obtained by a stochastic integrate-and-fire model in which the membrane voltage is represented as a random walk with drift (Gerstein and Mandelbrot, 1964). This simplified model was first suggested in (Schroedinger, 1915), and first applied in spike train data analysis by Gerstein and Mandelbrot (Gerstein and Mandelbrot, 1964).

(3) The probability distribution of the interspike intervals generated by a leaky integrate-and-fire model with Gaussian input can be derived exactly under the condition that the input is balanced (Feng and Ding, 2004).

Its $\delta\ \varepsilon\ \nu\ \sigma\ \iota\ \tau\ \psi$ is given by

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

where γ is the membrane time constant, μ and σ are the mean and the variance of the input. The previous distribution can be generalized as

$$p_{\mathrm{IF}}(t; \gamma, \mu, \sigma, \tau) = \begin{cases} f_{\mathrm{IF}}(t-\tau; \gamma, \mu, \sigma) & \text{if } t > \tau \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

in order to include a refractory period τ in the model.

(4) Finally, the log-normal probability density with parameters μ and σ

$$p_{\mathrm{LN}}(t; \mu, \sigma) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log t - \mu)^2}{2\sigma^2}\right]$$

is a suitable model when the log of the ISI are approximately normally distributed, see e.g. (Bhumbra and Dyball, 2004; McKeegan, 2002).

2.2. Parameter estimation for mixture models, the EM algorithm

Given the model (1) and a set of observations $\{t_1, t_2, \ldots, t_N\}$, the maximum-likelihood estimate of the parameter Θ, is that which maximizes the likelihood function (Papangelou, 1972; Berman, 1983; Tanner, 1996; Pawitan, 2001),

$$L(\Theta) = \prod_{n=1}^{N} p(t_n|\Theta) \tag{5}$$

or, equivalently, the log-likelihood

$$\ell(\Theta) = \sum_{n=1}^{N} \log\ p(t_n|\Theta) \tag{6}$$

To find the maximum-likelihood estimate above, we use the EM algorithm (Dempster and Laird, 1977), which is summarized and re-derived to our cases below.

First, we introduce the *complete* data set $\{(t_i, z_i)$: $i = 1, \ldots, N\}$, where the values taken by the variable z are not observable, or missing. It is easily seen that (1) is the marginal distribution of t of the joint distribution of $y = (t, z)$ when

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

and

$$p(t|z, \Theta) = p_z(t|\theta_z)$$

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, the expected complete data log-likelihood is

$$Q(\Theta|\Theta^{(c)}) = \sum_{i=1}^{N} \left(\sum_{k=1}^{K} \hat{\alpha}_{i,k}\ \log\ w_k \right) + \sum_{i=1}^{N} \left(\sum_{k=1}^{K} \hat{\alpha}_{i,k}\ \log\ p_k(t_i|\theta_k) \right) \tag{7}$$

where

$$\hat{\alpha}_{i,k} = \frac{w_k^{(c)} p_k(t_i|\theta_k^{(c)})}{\sum_{j=1}^{K} w_j^{(c)} p_j(t_i|\theta_j^{(c)})} \quad k \in \{1, \ldots, K\}, \quad i \in \{1, \ldots, N\} \tag{8}$$

is the probability that the ith observation was generated by the kth component of the mixture, given a model of parameter $\Theta^{(c)}$.

The EM algorithm iteratively maximizes the function $Q(\Theta|\Theta^{(c)})$. More specifically, each iteration of the EM algorithm 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

$$\hat{w}_k = \frac{1}{N} \sum_{i=1}^{N} \hat{\alpha}_{i,k}, \quad k = 1, \ldots, K, \tag{9}$$

and

$$\hat{\theta} = \text{argmax}_\theta \sum_{i=1}^{N} \sum_{k=1}^{K} \hat{\alpha}_{i,k}\ \log\ p_k(t_i|\theta_k), \tag{10}$$

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

When $p_k\ (t|\theta_k)$ belongs to the exponential family, e.g. for the Gaussian and the exponential distribution, Eq. (10) usually has a closed-form solution.

For the Gamma distribution, we have

$$\begin{cases} a_j^{(c+1)} b_j^{(c+1)} = \langle t \rangle_j^{(c)} \\ \Psi(a_j^{(c+1)}) - \log b_j^{(c+1)} = \langle \log t \rangle_j^{(c)} \end{cases} \tag{11}$$

where ψ is the digamma function (Tuckwell, 1988). For the inverse Gaussian distribution, we have[1]

$$\begin{cases} \mu_j^{(c+1)} = \langle t \rangle_j^{(c)} \\ \dfrac{1}{\lambda_j^{(c+1)}} + \dfrac{1}{\mu_j^{(c+1)}} = \left\langle \dfrac{1}{t} \right\rangle_j^{(c)} \end{cases} \tag{12}$$

where we defined, for a function F,

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

2.3. Parameter estimation with censored data

Let us consider a multiple-trial data set, where the spike times are recorded during a series of trials of length T_{obs}. From the observed spike times we can define two sets of interspike intervals: the *regular* (or complete) intervals $\{t_i\}_{i \in R}$, for which both start and end spike times exist, and the *truncated* intervals $\{t_i\}_{i \in T}$, for which the missing end times are replaced by T_{obs}. This is illustrated in Fig. 1.

Assuming that the probability density of the interspike intervals in the absence of censoring, i.e. for trials of infinite length, is given by $p(t|\theta)$, the probability density of the regular intervals is easily calculated as

$$\tilde{p}(t|\theta) = \begin{cases} \dfrac{p(t|\theta)}{1 - S(T_{\text{obs}}|\theta)} & if\ t < T_{\text{obs}} \\ 0 & otherwise \end{cases} \tag{13}$$

[1] We used here the known property of the inverse Gaussian distribution, that, for an inverse Gaussian sample $(X_1, X_2, \ldots, X_n)$, the variables $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ and $V = \frac{\left\{ \sum_{i=1}^{n} (1/X_i - 1/\bar{X}) \right\}}{n}$ are independently distributed, and $n\lambda V$ is distributed as χ^2_{n-1} (see Chikkara and Folks, 1989).

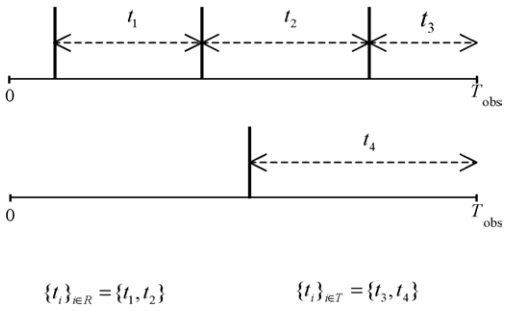

Fig. 1. Regular $\{t_i\}_{i\in R}$ and truncated $\{t_i\}_{i\in T}$ interspike intervals extracted from a multiple-trial spike data set.

where

$$S(t|\theta) = \int_t^{+\infty} p(t'|\theta)\,\mathrm{d}t'$$

A maximum-likelihood estimate for $\theta$is thus obtained by maximizing the function

$$L(\theta) = \sum_{i\in R} \log\ \tilde{p}(t_i|\theta) = \sum_{i\in R} \log\ p(t_i|\theta) - N_{\mathrm{R}} \log\,[1 - S(T_{\mathrm{obs}}|\theta)]$$

where N_{R} is the number of regular intervals observed. Obviously, this estimate becomes widely inaccurate for small N_{R}. A way to improve the accuracy is to take into account also the truncated intervals in the calculation of the likelihood. In particular, following (Cox and Lewis, 1966), we can use the following expression of the log-likelihood

$$L(\theta) = \sum_{i\in R} \log\ p(t_i|\theta) + \sum_{i\in T} \log\ S(t_i|\theta) \tag{14}$$

When p is given by a mixture of distributions, the log-likelihood of the complete data set $\{(t_i,\ z_i)\}_{i\in R,T}$ can be written as

$$L(\theta) = \sum_{i\in R} \log\ w_{z_i} + \sum_{i\in R} \log\ p_{z_i}(t_i|\theta_{z_i}) + \sum_{i\in T} \log\ w_{z_i} + \sum_{i\in T} \log\ S_{z_i}(t_i|\theta_{z_i}) \tag{15}$$

In order to compute the expectation of $L(\theta)$, we need the conditional probability density for the missing variable on the condition of the observed 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, this is again given by Eq. (8). For truncated intervals, instead, we cannot use (8), since the exact value of the interval is unknown. However, from the Bayes theorem we get

$$\hat{\alpha}_{i,k} = \frac{w_k^{(c)} S_k(t_i|\theta_k^{(c)})}{\sum_{j=1}^{K} w_j^{(c)} S_j(t_i|\theta_j^{(c)})} \quad \text{when } i \in T \tag{16}$$

Therefore, in presence of censored observations, Eq. (7) must be rewritten as

$$Q(\Theta|\Theta^{(c)}) = \sum_{i\in R}\sum_{k=1}^{K} \hat{\alpha}_{i,k}\ \log\ w_k + \sum_{i\in R}\sum_{k=1}^{K} \hat{\alpha}_{i,k}\ \log\ p_k(t_i|\theta_k) + \sum_{i\in T}\sum_{k=1}^{K} \hat{\alpha}_{i,k}\ \log\ w_k + \sum_{i\in T}\sum_{k=1}^{K} \hat{\alpha}_{i,k}\ \log\ S_k(t_i|\theta_k) \tag{17}$$

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

$$\hat{w}_k = \frac{1}{N}\left(\sum_{i\in R} \hat{\alpha}_{i,k} + \sum_{i\in T} \hat{\alpha}_{i,k}\right), \quad k = 1, \ldots, K, \tag{18}$$

$$\hat{\theta} = \mathrm{argmax}_\theta \left(\sum_{i\in R}\sum_{k=1}^{K} \hat{\alpha}_{i,k}\ \log\ p_k(t_i|\theta_k)\ s + \sum_{i\in T}\sum_{k=1}^{K} \hat{\alpha}_{i,k}\ \log\ S_k(t_i|\theta_k)\right) \tag{19}$$

Although a closed form for the solution of (19) does not exist in the general case, the problem can still be solved numerically with relative ease, given the independence of the parameters of different modules. A MATLAB implementation of the algorithm is available for download from http://www.informatics.sussex.ac.uk/users/er28/em/.

3. Results

3.1. Experimental data

As a first example, we consider a data set composed of 975 interspike intervals recorded in vitro from a goldfish retinal ganglion cell (Iyengar and Liao, 1997). Recordings are made with an extracellular microelectrode under constant illumination. As shown in Fig. 2 (top), the spike train is highly irregular, with brief bursts of spikes alternating to periods of silence. This is also reflected in the apparently bimodal character of the ISIH. The data are fitted by mixtures of two and three components, using the EM algorithm described in Section 2.2. The mixture weights are initialized randomly according to a uniform distribution. To initialize the parameters of the mixture components, we first group the data into n clusters using the K-means clustering algorithm, with n the number of components of the mixture. Then, we sort the clusters in ascending order by their centres, and finally fit each

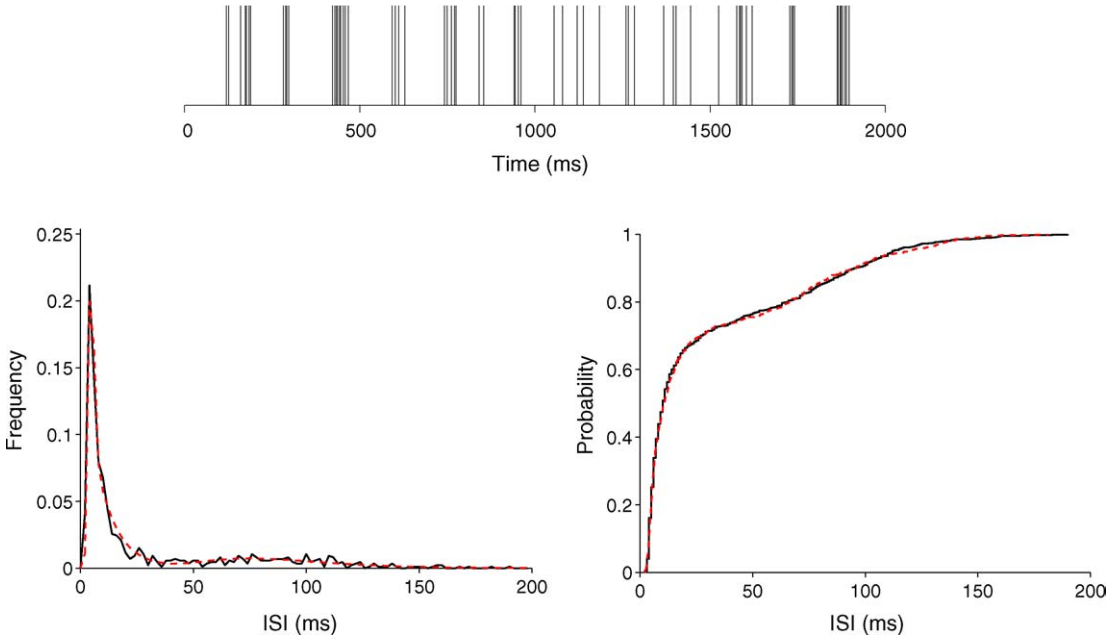

Fig. 2. Fitting retinal neuron spike data by mixture models. (Top) Raster plot of two seconds of spike times from a retinal ganglion neuron recorded in vitro under constant illumination. (Bottom left) Comparison between the experimental ISI distribution (solid line), and the distribution of the intervals generated by the three-component mixture model with the highest *P*-value (dashed line). (Bottom right) Comparison of the empirical cumulative distribution functions for experimental data (solid line), and model (dashed line).

mixture component to the corresponding cluster. The values of the estimated parameters are chosen as initial guesses for the EM algorithm. Goodness-of-fit is assessed by a two-sided two-sample KS test, where the data set used for fitting is compared with a sample drawn from the distribution of the estimated model.

As shown in Table 1, all the two-component mixtures considered fail to pass the KS test ($P < 0.05$). Goodness-of-fit is improved when using three-component mixtures and some of the estimated mixture models meet the criterion for acceptance.

Table 1
Fitting retinal neuron spike data by mixture models

Model	*P*	KS
(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

The *P*-values and KS statistics obtained for the estimated mixture models (1 = inverse Gaussian, 2 = Gamma). The model with the highest *P*-value is shown in bold.

The best-fit is achieved by the model

$$\hat{p}(t) = w_1 p_{\mathrm{Gam}}(t; a, b) + w_2 p_{\mathrm{IG}}(t; \mu_2, \lambda_2) + w_1 p_{\mathrm{IG}}(t; \mu_3, \lambda_3) \quad (20)$$

of parameters $(w_1, w_2, w_3) = (0.2592,\ 0.4912,\ 0.2497)$, $(a, b) = (15.1673,\ 0.3243)$, $(\mu_2,\ \lambda_2) = (13.6612,\ 24.9184)$, and $(\mu_2, \lambda_2) = (90.9478,\ 750.7258)$.

In Fig. 2 (left), we compare the experimental ISIH with the histogram of the intervals generated by model (20). The empirical cumulative distribution functions for the experimental data and the model are compared in Fig. 2 (right).

In order to verify that the method provides consistent indications when applied to data collected in homogeneous conditions, we consider the distribution of the parameters estimated by fitting the same mixture model to a set of 1000 independent samples, each one containing 1000 interspike intervals drawn from the distribution (20). The model used for fitting is a three-component mixture with the same components as (20).

KS tests performed on estimated models are positive for 95% of 1000 trials.

Fig. 3 shows the histograms of the weights, means and coefficients of variation for each the components of the estimated mixture models. Although the results indicate a considerable variability for some of the parameters, due to the partial overlap between the distributions, the method proves to be successful in identifying the underlying model features, even from samples of a limited size.

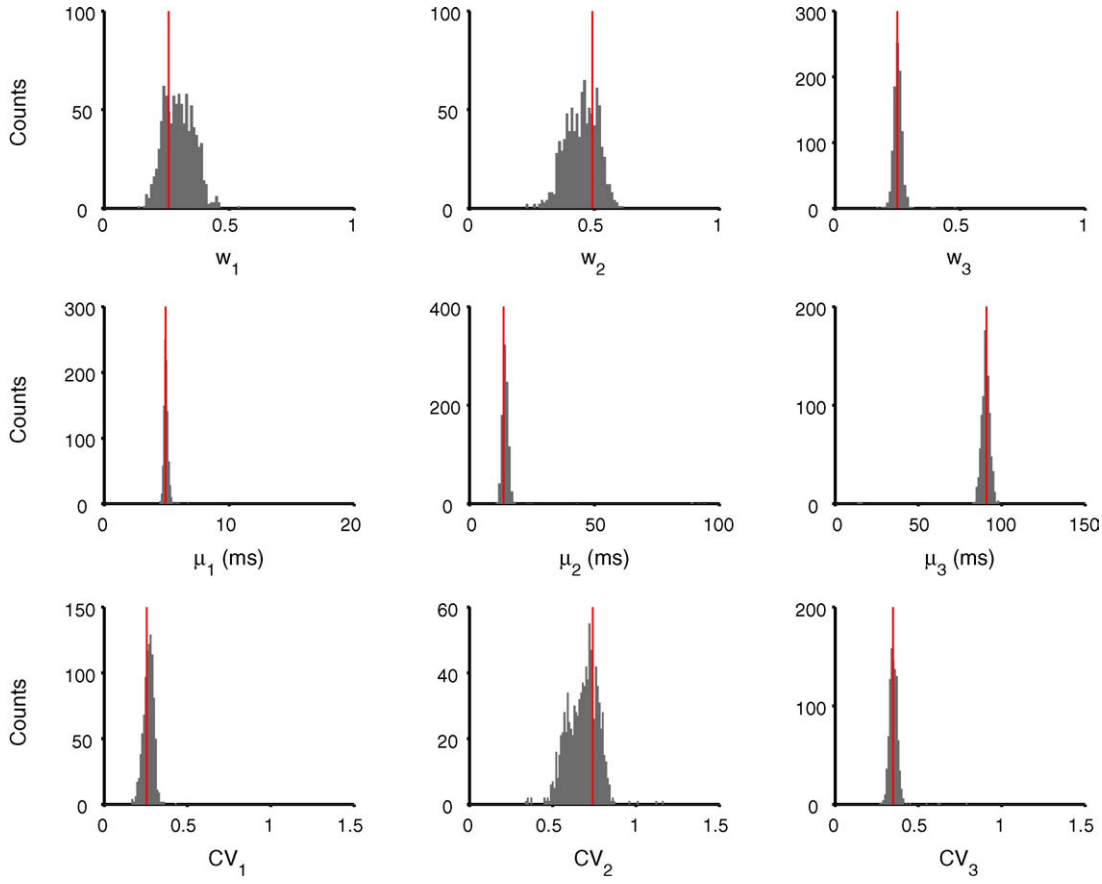

Fig. 3. Histograms of the estimated parameters for a three-component mixture model. The estimates were obtained by applying the EM algorithm on a series of 1000 independent trials. For each trial a sample of 1000 intervals drawn from the distribution (20) was considered. Vertical red lines indicate the true values of the parameters. Note that, for reason of clarity, we have reported here the distribution of the mean (μ) and the coefficient of variation (CV) of each component, instead of the distribution of the original parameters. For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.

3.2. Simulation data

To further test our approach, we consider a data set of 10^4 interspike intervals generated by numerical simulation of the Hodgkin–Huxley model[2],

$$C\frac{\mathrm{d}V}{\mathrm{d}t} = -g_{\mathrm{Na}}m^3h(V - V_{\mathrm{Na}}) - g_{\mathrm{K}}n^4(V - V_{\mathrm{K}}) - g_{\mathrm{l}}(V - V_{\mathrm{l}}) + I(t) \quad (21)$$

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{m_\infty(V) - m}{\tau_m(V)}, \quad \frac{\mathrm{d}h}{\mathrm{d}t} = \frac{h_\infty(V) - h}{\tau_h(V)}, \quad \frac{\mathrm{d}n}{\mathrm{d}t} = \frac{n_\infty(V) - n}{\tau_n(V)} \quad (22)$$

with input

$$I(t) = \mu + \sigma\xi(t)$$

where $\xi(t)$ is a Gaussian white noise, $\langle\xi(t)\rangle = 0$, $\langle\xi(t)\xi(t')\rangle = \delta(t - t')$. Values used for simulation were $\mu = 7$, $\sigma = 2$.

In the absence of noise, the system is characterized by the co-presence of two attractors, a fixed point and a limit cycle, which correspond respectively to a depolarized resting state, and to repetitive firing. For $\sigma > 0$, the solution trajectories hop randomly between these two attractors, thereby generating a burst-like pattern of activity, as seen in Fig. 4. It is reasonable to expect that the ISI distribution could be described by a mixture of distributions, rather than by a single probability density function.

The data are fitted with mixtures of two and three components, as described in the previous example. The obtained *P*-values and KS statistics for all the estimated models are reported in Table 2. Again the best-fit is achieved by a mixture of one Gamma and two inverse Gaussian distributions,

$$\hat{p}(t) = w_1 p_{\mathrm{Gam}}(t; a, b) + w_2 p_{\mathrm{IG}}(t; \mu_2, \lambda_2) + w_1 p_{\mathrm{IG}}(t; \mu_3, \lambda_3) \quad (23)$$

where the estimated parameters are $(w_1, w_2, w_3) = (0.3761, 0.4091, 0.2148)$, $(a, b) = (126.673, 0.1)$, $(\mu_2, \lambda_2) = (4477.1, 15.1)$ and $(\mu_3, \lambda_3) = (247.9, 28.8)$.

[2] For a detailed definition of the model, we refer the reader to Hodgkin and Huxley (1952).

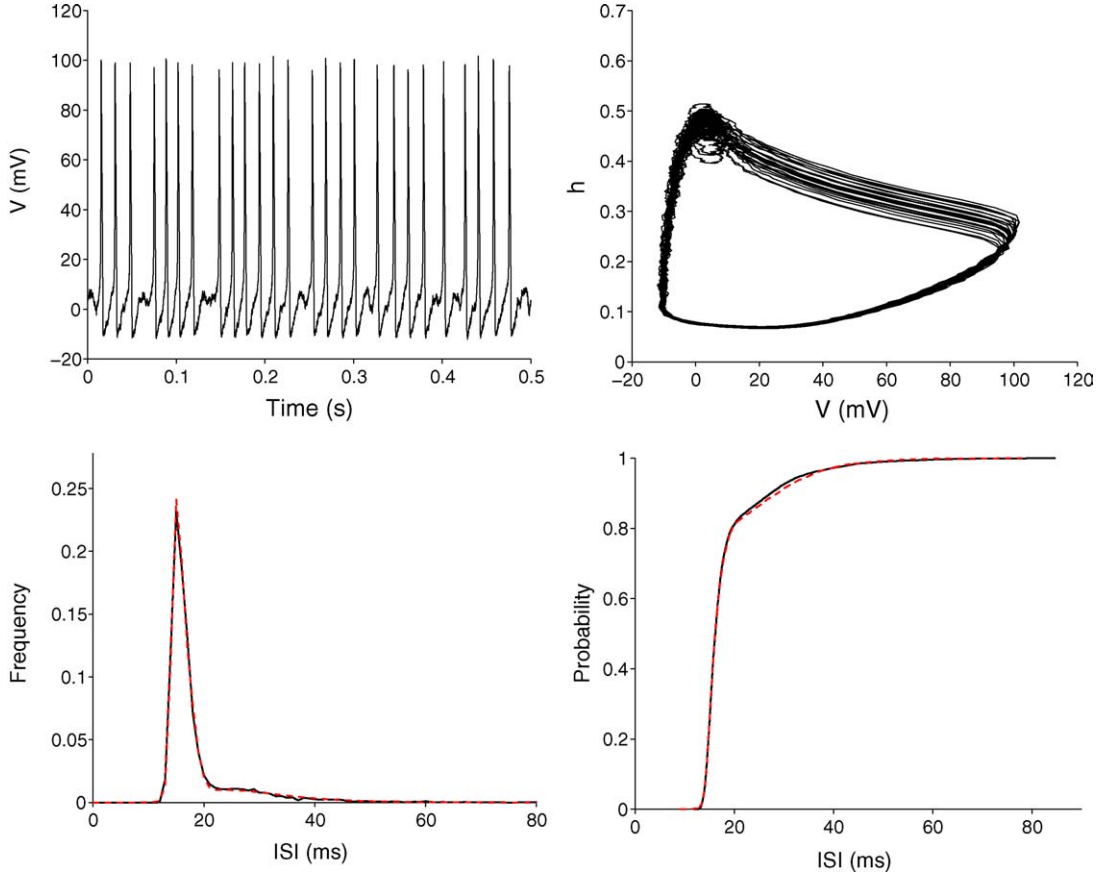

Fig. 4. Fitting simulated spike data by mixture models. (Top left) Time course of the membrane potential during a simulation run of models (21) and (22). (Top right) The phase trajectory of the variables *(V, h)* during the same simulation run. (Bottom left) Comparison between the distribution of the ISI generated by the HH model (solid line), and by the three-component mixture model with the highest *P*-value (dashed line). (Bottom right) Comparison of the empirical cumulative distribution functions for the HH model (solid line) and mixture model (dashed line).

Fig. 4 shows the compared histograms of the interspike intervals generated by the biophysical model, and by the mixture model (23), together with the respective empirical cumulative distribution functions.

Table 2
Fitting a simulated spike data by mixture models

Model	*P*	KS
(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
(3,4)	0.20	0.017

The *P*-values and KS statistics obtained for the estimated mixture models (1 = inverse Gaussian, 2 = Gamma, 3 = log-normal, 4 = integrate-and-fire). The models with the highest *P*-value are shown in bold.

It should be noted that, in theory, any distribution function can be fitted arbitrarily well by a mixture of sufficiently many components. However, increasing the number of components also results in models of increasing complexity, and makes parameter estimation less reliable, especially from small samples. Therefore, in selecting a model, one should always aim to a trade-off between accuracy and complexity.

A typical approach to model selection is represented by the Akike's Information Criterion (Pawitan, 2001). In this case, we ask whether the model could not be simplified by varying the distribution types used in the mixture, based on our understanding of the underlying system dynamics. As noted above, the phase trajectories switch randomly between the attractive basins of the two attractors. In particular, each time the phase trajectory enters the attractive basin of the limit cycle, the system starts firing almost periodically. Eventually, the system stops firing when a perturbation pushes the trajectory back into the attractive basin of the resting state. This phenomenon is clearly illustrated in Fig. 4, where we show a phase trajectory in terms of the variables V and h.

During phases of repetitive firing, it is reasonable to assume that the interspike intervals will be normally distributed around the unperturbed firing period, especially for

small σ. However, since the interspike intervals are always positive, a log-normal may represent a more adequate approximation. On the other hand, the escape times from the attractors, which correspond to the inter-burst intervals, are exponentially distributed in the limit $\sigma \to 0$, as suggested by the large deviation theory (Freidlin and Wentzell, 1984). As σ is increased, however, we expect that (4) may provide a better approximation of the ISI distribution, especially for short intervals.

A mixture including these two distributions is therefore a good candidate to approximate the distribution of the ISI produced by the model. Indeed the goodness-of-fit obtained after fitting the model to the data, is comparable with what obtained with the three-component mixture (23), see Table 2 (bottom line).

That the fitting procedure, as assessed by the KS test here, results in an acceptable *P*-value, is obviously not decisive in itself. Instead, as we demonstrated above, it is opportune that the choice of the distributions included in the mixture be dictated by functional interpretations, or by working hypotheses.

3.3. Censored data

Finally, we illustrate two examples of the application of our approach in the presence of censored data sets.

3.3.1. Fitting ISI distribution from multiple-trial spike data

A simulated multiple-trial data set is constructed as follows. First, 3000 intervals are randomly generated according to the distribution

$$p(t|\Theta) = \sum_{i=1}^{3} w_i p_{\mathrm{Gam}}(t|a_i, b_i), \tag{24}$$

of parameters $(w_1, w_2, w_3) = (0.55,\ 0.11,\ 0.34)$, $(a_1,\ a_2,\ a_3) = (25.43,\ 54.40,\ 5.21)$, and $(b_1,\ b_2,\ b_3) = (5.37,\ 2.92,\ 84.04)$. The spike train obtained by cumulating the generated intervals, is then split into 1438 consecutive chunks of length $T_{\mathrm{obs}} = 500$ ms, which are assumed to represent an ensemble of independent trials. Note that, due to severe censoring, about 25% of the simulated trials contain only one spike, as illustrated in the raster plot of Fig. 5. Finally, regular (1676) and truncated (1324) interspike intervals are extracted from all trials, as explained in Section 2.3.

The data are fitted by a three-component mixture model with the same composition as Eq. (24), using two different methods for parameter estimation. In the first case, the data set used for estimation includes only the regular intervals. In the second case, we consider both regular and truncated intervals, and apply the *censored* EM algorithm discussed in Section 2.3. The parameters of the distribution (24) and of the two estimated mixture models are compared in Table 3.

From the comparison of the ISI histograms, Fig. 5, it is particularly evident that the model estimated from the regular intervals provides only a poor approximation of the true, i.e. uncensored, ISI distribution. On the other hand, the result improves significantly after considering also the truncated intervals in the estimation.

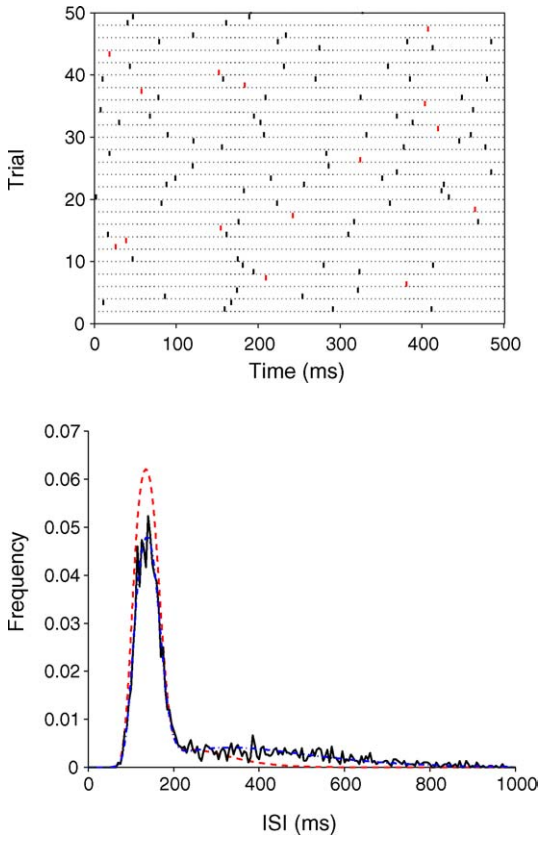

Fig. 5. Fitting multiple-trial data with mixture models. (Top) Spike raster plots of the first 50 simulated trials. Trials containing a single spike (not a complete interspike interval) have spikes drawn in red; horizontal dotted lines have been added for visual clarity. (Bottom) The distribution of the uncensored interspike intervals (black solid line), compared with the distribution of intervals generated by the model estimated by the standard (red dashed line), and the *censored* EM algorithm (blue dot-dashed line). For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.

3.3.2. Decoding neural activity from multielectrode array data

We consider a group of N leaky integrate-and-fire neurons. Given a time-dependent signal $\lambda(t)$, the input to the ith neuron

Table 3
Fitting simulated multiple-trial spike data with mixture models

Model	w_1	μ_1	σ_1	w_2	μ_2	σ_2	w_3	μ_3	σ_3
p	0.55	137	27	0.11	156	21	0.34	438	192
$\hat{p}_{\mathrm{EM}}$	0.62	128	22	0.22	159	17	0.16	261	88
$\hat{p}_{\mathrm{CEM}}$	0.35	125	20	0.30	157	20	0.35	413	185

The parameters of the original mixture model (p), and of the mixture models estimated using the standard ($\hat{p}_{\mathrm{EM}}$) and the censored ($\hat{p}_{\mathrm{CEM}}$) EM algorithm. Note that, instead of the original parameters (a, b), we have reported here the mean (μ) and the standard deviation (σ) of each component.

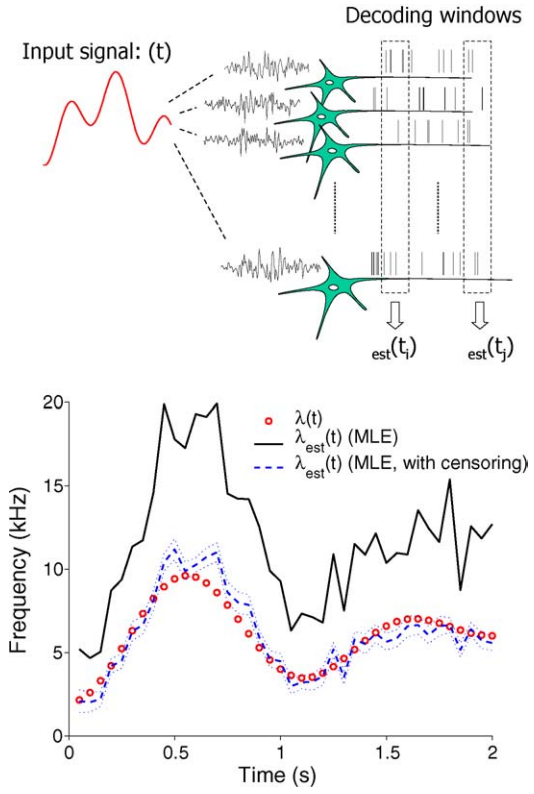

Fig. 6. Decoding neural activity from a spike train ensemble. (Top) Schematic drawing of the setup. (Bottom) Comparison between the input signal $\lambda(t)$ (circles), the signal estimated from the analysis of the regular interspike intervals (solid line), and the signal estimated from the regular and the truncated intervals (dashed). The dotted lines indicate 95% confidence intervals for the latter estimate. The spike activity was decoded over consecutive 50 ms stretches.

is given by

$$I_i(t) = \mu(t) + \sigma(t)\xi_i(t), \quad i = 1, \ldots, N$$

where $\xi_i(t)$ is a Gaussian white noise, with $\langle \xi_i(t) \rangle = \delta ij \delta(t - t')$ and the parameters of the input are

$$\mu(t) = a\lambda(t)(1 - r(t)), \quad \sigma(t) = a\sqrt{\lambda(t)(1 + r(t))} \tag{25}$$

with

$$r(t) = 1 - \frac{V_{\text{thre}}}{a\lambda(t)\gamma} \tag{26}$$

The last equation ensures that the input is *balanced* at all times (Feng and Ding, 2004).

For the sake of simplicity, we consider here a homogeneous population, so that all neurons conform to the same density function (4). However, this assumption is not crucial to applicability of the method (see Section 4), and could be relaxed. The parameters used in the simulation were $N = 100$, $\gamma = 20$ ms, $V_{\text{thre}} = 20$ mV, and $a = 0.5$ mV. The signal was given by $\lambda(t) = \lambda_0 + 2\lambda_0 \sin^2(\upsilon_1 t) + 2\lambda_0 \sin^2(\upsilon_2 t)$, with $\upsilon_1 = 3.14$ Hz, $\upsilon_2 = 2.36$ Hz, $\lambda_0 = 2$ kHz. The setup is illustrated schematically in Fig. 6.

For decoding, the spike trains are analyzed within consecutive 50 ms time windows. For each time window, regular and truncated interspike intervals are extracted from spike data, and the stimulus is estimated by fitting the theoretical distribution (3) to the set of regular intervals, or to the joint set of regular and truncated intervals.

Fig. 6 shows the original and the decoded signals. It is easily seen that taking into account the censoring in the estimate, the signal can be correctly retrieved even with such small decoding windows.

4. Discussion

We have presented a semi-parametric approach to fit interspike interval distributions, and assessed the potential of this approach on data from both experiments and simulations. In addition, we have addressed the issue of fitting censored data, corresponding to the case of multiple-trial and multi-electrode array data.

The decoding task in Section 3.3.2 is also of paramount importance for sensory information processing in the brain (Kamitani and Tong, 2005). Let us assume that the synaptic input to a group of downstream neurons encode the information about a time-dependent sensory stimulus. Given the neuronal output, in the form of the emitted spike trains, it is natural to ask how accurately an external observer, or some other neurons further downstream, can retrieve the original information. In other terms, it is important to know to what extent the original information is preserved in the output spike trains and is later accessible for further processing. For rapidly varying stimuli, in particular, the output has to be decoded quickly in order to produce a prompt response (Thorpe et al., 1996). However, when the population activity is sampled over such short time windows, the censoring effect becomes prominent, and it may be questioned to what extent it is still possible to correctly retrieve the stimulus. To address this issue, we have implemented an *optimal* decoding strategy based on 'censored' maximum likelihood estimates. The results obtained for this ideal case have more than just a theoretical interest; they could also be used to provide an upper bound (ideal observer) for the accuracy with which any other method can retrieve or process the original information (Deneve et al., 1999).

4.1. Why fitting?

It is certainly unsatisfactory to approximate or to fit a distribution, rather than to give an accurate description of it with analytical expressions. However, we know from our past years of experience that to derive exact probability distributions for the interspike intervals of realistic neuronal models is a formidable challenge, and in fact, this problem is already particularly difficult even for the simple leaky integrate-and-fire model. Moreover, we do not expect to have any analytical expression that can represent the interspike interval distribu-

tion of a cortical cell in vivo. Hence, the use of approximated probabilistic models and of fitting methods to represent the properties of the interspike intervals is an efficient way to describe the neuronal dynamics.

4.2. Dimensionality reduction

With multi-electrode arrays it is now possible to record spatio-temporal patterns of activity in large populations of neurons. However, these patterns are in a high dimensional space and it is usually difficult to grasp their meaning. By fitting known probability distributions to the data, we can greatly reduce the dimensionality of the problem, and we are in a much better position to extract information from these sort of data. An example of this approach is shown in Section 3.3.2. Of course, we would expect the reduced dimensions to be in general more than one, for example including the weights and the parameters of each distribution in a mixture.

However, care must be taken when interpreting the estimated parameters of a mixture model in physiological terms, since the latter are usually not unique, as shown before. This issue, which is common to many complex optimization problems, is related to the existence of local extrema in the functions that have to be minimized/maximized, and is required further considerations in the future.

4.3. Decoding

One of the most challenging tasks in neuroscience is to understand how information is encoded in an ensemble of spike trains, and how it can be decoded from them. We have demonstrated how our approach can be applied to such problems. These preliminary results are encouraging and an application to multi-electrode array data from the olfactory bulb (Horton et al., 2005) will be presented shortly.

4.4. Temporal information

Finally, it can be argued that by fitting interspike intervals with probability distributions, any temporal information about the stimulus is lost. However, this is not true if we look at the problem from a population coding point of view. As mentioned above, we can pool together the spike trains generated by a population of neurons, and analyze them within moving time windows. In each time window, we can fit our model to the data to obtain a set of parameters describing the stimulus. Then the whole information stored in the population spike trains will be summarized in the temporal dynamics of the parameters, and the information in the time domain can be fully recovered.

In the current paper, we have treated each decoding window as independent (see Fig. 6). Of course we can easily include temporal relationship, as we mentioned above. Another way to include temporal relationship is to introduce nonhomogeneous point processes, as developed, e.g. in (Brown et al., 2004). It would be interesting to see how the current results can be improved if we combine two approaches together, namely decoding with a nonhomogeneous point process using a mixture of distributions as its interspike interval distribution.

4.5. Homogeneity versus heterogeneity in decoding

In Section 3.3.2, we have considered how to decode information from a group of homogeneous neurons (integrate-and-fire models). However, it is natural to ask whether our method is also applicable to heterogeneous populations of neurons that are typically observed in experiments.

Let us consider a set of N neurons having different thresholds $V_{\text{thre}}^{(i)}$. Then a maximum-likelihood estimate of the stimulus can be written in the following form

$$\hat{\lambda} = \sum_{i=1}^{N} \frac{f(V_{\text{thre}}^{(i)}, \{T_j^{(i)}\})}{N}$$

where f is some function, and $\{T_j^{(i)}\}$is the spike train emitted from the ith neuron. With the mean-field approximation, we have

$$\hat{\lambda} = \sum_{i=1}^{N} \frac{f(\langle V_{\text{thre}}^{(i)} \rangle, \{T_j^{(i)}\})}{N}$$

where $\langle V_{\text{thre}}^{(i)} \rangle$ is the average over the population of neurons. Thus, in this scenario all the conclusions derived in the current paper are true. Further investigation as to what extent such a conclusion holds true, is surely worth and it will constitute our future topic of research.

Acknowledgment

We would like to thank E.N. Brown for providing us with the goldfish data, and for useful discussion about censoring. J.F. was partially supported by grants from UK EPSRC(GR/R54569), (GR/S20574), and (GR/S30443).

References

Andersen PK, Borgan O, Gill RD, Keiding N. Statistical models based on counting processes. New York: Springer-Verlag; 1993.

Barbieri R, Quirk MC, Frank LM, Wilson MA, Brown EN. Construction and analysis on non-Poisson stimulus-response models of neural spike train activity. J Neurosci Methods 2001;105:25–37.

Berman M. Comment on 'Likelihood analysis of point processes and its applications to seismological data' by Y. Ogata. B Int Statist Inst 1983;50:412–8.

Bhumbra GS, Dyball EJ. Measuring spike coding in the rat supraoptic nucleus. J Physiol 2004;555:281–96.

Brillinger DR. Maximum likelihood analysis of spike trains of interacting nerve cells. Biol Cybern 1988;59:189–200.

Brown EN, Barbieri R, Ventura V, Kass RE, Frank LM. The time-rescaling theorem and its application to neural spike train data analysis. Neural Comput 2002;14:325–46.

Brown EN, Frank LM, Tang D, Quirk MC, Wilson MA. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. J Neurosci 1998;18:7425–8411.

Brown EN, Nguyen DP, Frank LM, Wilson MA, Solo V. An analysis of neural receptive field plasticity by point process adaptive filtering. P Natl Acad Sci USA 2001;98:12261–6.

Brown EN, Barbieri R, Eden UT, Frank LM. Likelihood methods for neural spike train data analysis. In: Feng JF, editor. Computational neuroscience: a comprehensive approach. Boca Raton: Chapman & Hall/CRC; 2004. p. 253–86.

Casella G, Berger RL. Statistical inference. Belmont, CA: Duxbury; 1990.

Chikkara RS, Folks JL. The inverse Gaussian distribution: theory, methodology, and applications. New York: Marcel Dekker; 1989.

Cox PR, Lewis PA. The statistical analysis of series of events. New York: Methuen; 1966.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser A-G 1977;39: 1–38.

Deneve S, Latham P, Pouget P. Reading population codes: a neural implementation of ideal observers. Nat Neurosci 1999;2:740–5.

Feng J, Ding M. Decoding spikes in a spiking neuronal network. J Phys A: Math Gen 2004;37:5713–27.

Frank LM, Brown EN, Wilson MA. Trajectory encoding in the hippocampus and entorhinal cortex. Neuron 2000;27:169–78.

Frank LM, Eden UT, Solo V, Wilson MA, Brown EN. Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. J Neurosci 2002;22:3817–30.

Freidlin MI, Wentzell AD. Random perturbation of dynamical systems. New York: Springer-Verlag; 1984.

Gerstein GL, Mandelbrot B. Random walk models for the spike activity of a single neuron. J Biophys 1964;4:41–68.

Guttorp P. Stochastic modeling of scientific data. London: Chapman & Hall; 1995.

Hodgkin A, Huxley A. A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 1952;117:500–44.

Horton PM, Bonny L, Nicol AU, Kendrick KM, Feng JF. Applictions of multi-variate analysis of variances (MANOVA) to multi-electrode array data. J Neurosci Methods 2005;146:22–41.

Iyengar S, Liao Q. Modeling neural activity using the generalized inverse Gaussian distribution. Biol Cybern 1997;77:289–95.

Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. Nat Neurosci 2005;8:679–85.

Kass RE, Ventura V. A spike train probability model. Neural Comput 2001;13:1713–20.

Levine MW. The distribution of intervals between neural impulses in the maintained discharges of retinal ganglion cells. Biol Cybern 1991;65:459–67.

McKeegan DEF. Spontaneous and odour evoked activity in single avian olfactory bulb neurones. Brain Res 2002;929:48–58.

McLachlan GJ, Peel D. Finite mixture models. New York: Wiley; 2000.

Papangelou F. Integrability of expected increments of point processes and a related random change of scale. Trans Am Math Soc 1972;165:483–506.

Pawitan Y. In all likelihood: statistical modelling and inference using likelihood. London: Oxford; 2001.

Schroedinger E. Zur Theorie der fall- und steigversuche an teilchen mit Brownscher bewegung. Phys Ze 1915;16:289–95.

Snyder D, Miller M. Random point processes in time and space. 2nd ed. New York: Springer-Verlag; 1991.

Tanner MA. Tools for statistical inference: methods for the exploration of posterior distribution and likelihood functions. New York: Springer-Verlag; 1996.

Taylor HM, Karlin S. An introduction to stochastic modeling. San Diego CA: Academic Press; 1994.

Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature 1996;381:520–2.

Tuckwell HC. Introduction to theoretical neurobiology: nonlinear and stochastic theories. New York: Cambridge; 1988.