Formatting Instructions for NIPS 2013

Anonymous Author(s)
Affiliation
Address
email

Abstract

1 The model/introduction

Our data is a time-series of multielectrode recordings $\mathbf{X} \equiv (\mathbf{x}_1, \cdots, \mathbf{x}_T)$, and consists of T recordings from C channels. The set of recording times lie on regular grid with interval length Δ, while $\mathbf{x}_t \in \mathbb{R}^C$ for all t. This time-series of electrical activity is driven by an unknown number of neurons and we want to... outline scientific goals. We let the number of neurons be unbounded, though only a few of the infinite neurons dominate. These neurons contribute the majority of the activity in any finite interval of time; however, as time passes, the total number of observed neurons increases (Justify?). The neurons themselves emit continuous-time voltage traces, with the outputs of all neurons superimposed and discretely sampled to produce the recordings $\mathbf{X}$. At a high level, we model the output of each neuron as a series of idealized spikes smoothed with appropriate kernels (the latter determines the shape of each action potential). We describe this in detail, starting first with the model for a single channel recording $X \equiv (x_1, \cdots, x_T)$.

1.1 Modelling a single neuron output

We model the spiking activity of each neuron as stationary and memoryless, with its set of spike times distributed as a homogeneous Poisson process. Comment on refractoriness or leave for discussion/future work. Similarly on the generalization to inhomogeneity? The neurons themselves are heterogeneous, with r_i the (unknown) firing rate for neuron i. Call the ordered set of spike times of the ith neuron E_i; then the time between successive elements of E_i is exponentially distributed with mean $1/r_i$. We write this as

$$E_i \sim \text{PoissProc}(r_i) \tag{1}$$

The actual electrical output of a neuron is not a binary event; instead each spiking event is a smooth voltage perturbation about a resting state. This perturbation forms the shape of the spike (without any loss of generality, we set the resting state to zero). (figure? better biological description? comment on how we preprocess the data to get zero mean?). While the spike shape varies across neurons as well as across different spikes of the same neuron, each neuron has its own characteristic distribution over shapes. Figure? We let $\theta \in \Theta$ parametrize this distribution, and whenever neuron i emits a spike, we draw a voltage trace independently from the corresponding distribution. This is then offset to the time of the spike, and the complete output of the neuron is the superposition of all these spike waveforms. (Figure?) More concretely, we model each spike shape as a linear combination of a dictionary of K basis functions $A \equiv (A_1(t), \cdots, A_K(t))$, shared across all neurons. For the ith neuron, the jth spike $e_{ij} \in E_i$, is associated with a random K-dimensional weight vector $\tilde{y}_{ij}$, and the shape of this spike is given by the weighted sum $\sum_{k=1}^{K} \tilde{y}_{ijk} A_k(t)$. We let $\tilde{y}_{ij}$ be Gaussian distributed, with $\theta_i \equiv (\mu_i, \Sigma_i)$ determining its mean and variance. Then, at any time t, the output of

neuron i is

$$x_i(t) = \sum_{j=1}^{|E_i|} \sum_{k=1}^{K} \tilde{y}_{ijk} A_k(t - e_{ij}) \tag{2}$$

The total signal recorded $x(t)$ at any electrode is the superposition of the outputs of all neurons. Define $E = \cup_{i=1}^{\infty} E_i$ as the (ordered) superposition of the spike times of all neurons. Furthermore, let $n(j)$ be the neuron to which the jth element of E belongs, and let $p(j)$ index the position of the jth spike of E in the spike train $E_{n(i)}$ of neuron $n(i)$ (so that $e_j = e_{n(j)p(j)}$). Then, we have that

$$x(t) = \sum_{i=1}^{\infty} x_i(t) = \sum_{j=1}^{|E|} \sum_{k=1}^{K} y_{jk} A_k(t - e_j) \tag{3}$$

where

$$y_j \equiv \tilde{y}_{n(j)p(j)} \sim N(\mu_{n(j)}, \sigma_{n(j)}) \tag{4}$$

From the superposition property of the Poisson process [1], the overall spiking activity E is a Poisson process with rate $R = \sum_{i=1}^{\infty} r_i$. The signal $x(t)$ is a functional of a marked Poisson process, where the jth event is labelled by the neuron to which it is assigned ($n(j)$), and the shape of its spike waveform (y_j). From the properties of the Poisson process, we have that the marks $n(j)$ are i.i.d. distributed with $P(n(j) = i) = \frac{r_i}{R}$. Given $n(j)$, y_j is distributed as in equation 4.

1.2 Completely random measures (CRMs)

In this work, we take a nonparametric approach, letting the number of neurons be unbounded (so that $n(i) \in \{1, 2, \cdots\}$). Since only a finite number of spikes are observed in any finite interval, the total rate R must also be finite; moreover, as we described earlier, we want this to be dominated by a few r_i. A natural framework that captures these modelling requirements is that of completely random measures [2]. Completely random measures are stochastic processes that form flexible and convenient priors over infinite dimensional objects like probability distributions, hazard functions etc. These have been well studied in the Bayesian nonparametrics and machine learning community, and there exists a wealth of literature on their theoretical properties, as well as on computational approaches to posterior inference.

Recall that each neuron is characterized by a pair (r_i, θ_i); the former characterizes the distribution over spike times, and the latter over spike shapes. We map the infinite collection of pairs $\{(r_i, \theta_i)\}$ to an atomic measure on Θ:

$$R(\mathrm{d}\theta) = \sum_{i=1}^{\infty} r_i \delta_{\theta_i} \tag{5}$$

For any subset Θ of Θ, the measure $R(\Theta)$ equals $\sum_{\{i:\theta_i \in \Theta\}} r_i$. We allow $R(\cdot)$ to be random, modelling it as a realization of a completely random measure. Such a random measure has the property that for any two disjoint subsets Θ_1 and $\Theta_2 \in \Theta$, the measures $R(\Theta_1)$ and $R(\Theta_2)$ are independent. This distribution over measures is induced by a distribution over the infinite sequence of weights (the r_i's), and a distribution over the sequence of their locations (the θ_i's). For a CRM, the weights r_i form the jumps of a Lévy process [3], and their distribution is characterized by a Lévy intensity $\rho(r)$. The locations θ_i are drawn i.i.d. from a base probability measure $H(\theta)$; we let this be the conjugate normal-Wishart distribution. As it typical, we assume these to be independent (though this is not necessary). if there's space, I can elaborate on the construction of the CRM from its Levy measure, though this is not necessary

The CRM we choose is the Gamma process (ΓP); this has Lévy intensity $\rho(r) = r^{-1} \exp(-r\alpha)$. The Gamma process has the convenient property that the total mass $R \equiv R(\Theta) = \sum_{i=1}^{\infty} r_i$ is Gamma distributed (and thus conjugate to the Poisson process prior on E). The Gamma process is also closely connected with the Dirichlet process [4], which will prove useful later on. Other choices of the Lévy intensity can capture greater uncertainty in the number of neurons active in any finite

interval, power-law behaviour etc. In any case, our overall model is then:

$$R(\mathrm{d}\theta) \sim \Gamma\mathrm{P}(\alpha, H(\theta)) \tag{6}$$

$$E_i \sim \text{PoissProc}(r_i) \quad i \text{ in } 1, 2, \cdots \tag{7}$$

$$\tilde{y}_{ij} \sim N(\mu_i, \Sigma_i) \quad i, j \text{ in } 1, 2, \cdots \tag{8}$$

$$x_i(t) = \sum_{j=1}^{|E_i|} \tilde{y}_{ij} A_j(t - e_{ij}) \tag{9}$$

$$X = \sum_{i=1}^{\infty} x_i \tag{10}$$

It will be more convenient from the point of inference to work with the marked Poisson process representation of equations 3 and 4. The superposition process E is a rate R Poisson process, and under a Gamma process prior, R has a conjugate Gamma distribution with shape and scale parameters 1 and α respectively. As we saw, the labels $n(\cdot)$ assigning events to neurons are drawn i.i.d. from a normalized Gamma process $G(\mathrm{d}\theta)$:

$$G(\mathrm{d}\theta) = \frac{r_j}{R} \tag{11}$$

$G(\mathrm{d}\theta)$ is a random probability measure that belongs to a class called a normalized random measures [5]; for the Gamma process, this is a draw from the Dirichlet process. For the j spike, given its neuron assignment $n(i)$, its shape vector is drawn from a normal distribution with parameters $(\mu_{n(j)}, \Sigma_{n(j)})$. Thus the weight vectors are distributed according to a Dirichlet process mixture model, with the neurons forming clusters. This insight allows us to marginalize out the infinite-dimensional rate vector R, and assign spikes to neurons via the Chinese restaurant process (CRP). Under the CRP, the jth spike is assigned to one of the earlier neurons with probability proportional to the number of earlier spikes assigned to that neuron. It is assigned to a new neuron with probability α. Unlike most applications with observe the outputs of a CRP, we observe a functional of it. Furthermore, (again, for the Gamma process), the random probability measure G is independent of the total mass $R(\Theta)$. We thus have the following model equivalent to the one above:

$$R \sim \text{Gamma}(1, \alpha) \tag{12}$$

$$G(\mathrm{d}\theta) \sim \text{DP}(\alpha) \tag{13}$$

$$E \sim \text{PoissProc}(R) \tag{14}$$

$$n(j) \sim G, \quad j = 1, \cdots, |E| \tag{15}$$

$$y_e \sim N(\mu_{n(e)}, \Sigma_{n(e)}), \quad j = 1, \cdots, |E| \tag{16}$$

$$x(t) = \sum_{j=1}^{|E|} \sum_{k=1}^{K} y_{jk} A_k(t - e_j) \tag{17}$$

1.3 A discrete-time approximation

In the previous paragraphs, we described a continuous-time voltage output by a neuron. Our data on the other hand consists of recordings at a discrete set of times. While it is possible to make inferences about the continuous-time process that underlies these discrete recordings, in this work we restrict ourselves to discrete-time inferences. Towards this, we start by providing a discrete-time approximation to the model above. This follows easily from the marked Poisson process characterization of the model. Recall first the Bernoulli approximation to the Poisson process: a sample from a Poisson process with rate R can be approximated by discretizing time at a granularity Δ, and assigning each interval an event independently with probability $R\Delta$. This approximation becomes exact as Δ tends to 0.

This suggests the following approximation at a time resolution Δ. Draw the random Poisson process rate R drawn from a Gamma$(1, \alpha)$ distribution. Simultaneously, draw a random probability measure G from a Dirichlet process. Assign an event to an interval independently with probability $R\Delta$, and to each event, assign a random mark drawn from the DP. Given the marks, we can evaluate the recordings at each time.

1.4 Noise and nonstationarity

The signal recorded by an electrode is the neuron output corrupted by noise, we model this noise as independent of the signal, additive and Gaussian. However, rather than modelling the noise as independent across time bins, we model it as a first-order autoregressive process. This can capture effects like the movement of electrodes during the experiment. Furthermore, rather than keeping the cluster parameters fixed, we model these as AR processes as well, capturing the evolution of the neuron shape with time.

1.5 Modelling multielectrode recordings

2 Inference

We perform inference in an online manner [6]. As observations arrive, our inference algorithm decides whether or not a new spike is present, which neuron (cluster) to assign that spike to, as well as the shape of the spike waveform. On the other hand, our algorithm maintains a posterior distribution over the cluster parameters that characterize the distribution over shapes. Having identified the location and shape of spikes from earlier times, we subtract these from the observations treat the residual as an observation from a DP mixture model. The cluster assignment of earlier spikes determines the seating arrangement of customers in the Chinese restaurant associated with the DP. Given the corresponding distribution over parameters, $p(\theta)$, we decide whether there is an underlying spike, which cluster it is assigned to, and what the shape of that spike is. We simultaneously update the distribution over parameters of clusters. Assume each spike waveform spans W time intervals. Define the residual at time t as $X_t - \sum_{i=1}^{W} A$. At time t, let y_t represent the shape of the action potential. Letting $\tilde{x}_t$ be the observation at time t, we have

$$z_t \sim Bern(p) \tag{18}$$

if $z_t == 1$

$$\gamma_t|\gamma_{1:t-1} \sim CRP \tag{19}$$

$$\theta_t|\gamma_t = i \sim \mathcal{N}(\theta_{t-1}, \Sigma) \tag{20}$$

References

[1] J. F. C. Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1993. Oxford Science Publications.

[2] J.F.C. Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.

[3] Ken ti Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1990.

[4] Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230, 1973.

[5] L.F. James, A. Lijoi, and I. Pruenster. Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.*, 36:76–97, 2009.

[6] L. Wang and D.B. Dunson. Fast bayesian inference in dirichlet process mixture models. *Journal of Computational & Graphical Statistics*, 2009.