Formatting Instructions for NIPS 2013

Anonymous Author(s)
Affiliation
Address
email

Abstract

1 The model

Our data is a time-series of multielectrode recordings $\mathbf{X} \equiv (\mathbf{x}_1, \cdots, \mathbf{x}_T)$, consisting 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. We let this 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 sequences of action potentials which are superimposed to produce the recordings $\mathbf{X}$. At a high level, we model the output of each neuron as a series of idealized spikes convolved with an appropriate smoothing kernel (the latter determines the shape of each action potential). We describe this in detail, starting first with the model for a single channel.

The distribution over spike times for each neuron is modelled as a homogeneous Poisson process. Let r_i be the unkown firing rate for neuron i, then letting E_i be the ordered set of its spike times, 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 spike is characterized by a smooth rise and fall in voltage (Not just rise and fall, figure?). The shape of this waveform varies across neurons as well as across different spikes from the same neuron, however each neuron has its own characteristic distribution over shapes. We let $\theta \in \Theta$ parametrize this distribution, and whenever neuron i emits a spike, draw a new smoothing kernel independently from the corresponding distribution p_{θ_i}, and place it at the time of the spike.

We model each action potential shape as a linear combination of dictionary of K basis function $A \equiv (A_1, \cdots, A_K)$, 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 y_{ij}, and the shape of this spike is given by $\sum_{k=1}^{K} A y_{ij}$. We let y_{ij} be normally distributed, with $\theta_i \equiv (\mu_i, \Sigma_i)$ determining the mean and variance. Thus, at any time t, the output of neuron i is

$$x_i(t) = \sum_{j=1}^{|E_i|} K_{ij}(t - e_j) \tag{2}$$

The total signal recorded $x(t)$ at any electrode is the superposition of the outputs of all neurons. Defining $E = \cup_{i=1}^{\infty} E_i$ as the (ordered) superposition of the spike times of all neurons, and $n(j)$ as the neuron to which the jth element of E belongs to, we have:

$$x(t) = \sum_{i=1}^{\infty} x_i(t) = \sum_{j=1}^{|E|} K_{n(j)s(j)}(t) \tag{3}$$

From the superposition property of the Poisson process [1], the overall spiking activity E is a realization of a Poisson process with rate $R = \sum_{i=1}^{\infty} r_i$. 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 three modelling requirements is that of completely random measures [2]. We map the infinite collection of pairs $\{(r_i, \theta_i)\}$ to an atomic measure on Θ:

$$R(\mathrm{d}\theta) = \sum_{i=1}^{\infty} r_i \delta_{\theta_i} \tag{4}$$

For any subset Θ of Θ the measure $R(\Theta)$ equals $\sum_{\{i:\theta_i \in \Theta\}} r_i$. We allow this to be random, modelling $R(\cdot)$ as a realization of a completely random measure (CRM). This is a random measure where for any two disjoint subsets Θ_1 and $\Theta_2 \in \Theta$, the measures $R(\Theta_1)$ and $R(\Theta_2)$ are independent. The distribution over measures is induced by the distribution over the infinite sequence of weights (the r_i's), and the distribution over the sequence of their locations (the θ_i's). For a completely random measure, the weights r_i form the jumps of a Lévy process [3], and that their distribution is characterized by a 'Lévy intensity'. The locations θ_i are drawn i.i.d. from a base probability measure $H(\theta)$. Completely random measures have been well studied in the Bayesian nonparametrics community where they form flexible and convenient priors over various quantities. Different choices of the Lévy intensity can capture uncertainty in the number of neurons active in any finite interval, power law behaviour etc.

The CRM we choose is the Gamma process; this has Lévy intensity $r^{-1}\exp(-r\alpha)$. This 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.

The overall model is then:

$$R(\mathrm{d}\theta) \sim \Gamma\mathrm{P}(\alpha) \tag{5}$$

$$E_i \sim \text{PoissProc}(r_i) \quad i \text{ in } 1, 2, \cdots \tag{6}$$

$$x_i(t) = \sum_{j=1}^{|E_i|} A y_j \delta_{(t-e_j)} \tag{7}$$

$$X = \sum_{i=1}^{\infty} x_i \tag{8}$$

It will be more convenient to view the set of pairs $\{(e_{ij}, y_{ij})\}$ as a realization of a marked Poisson process in time. The overall process E is a rate R Poisson process, with each event assigned a label or mark indicating the neuron to which it is assigned (or equivalently, the parameter θ associated with that neuron). These marks are drawn from a probability measure $G(\mathrm{d}\theta) = \frac{1}{R}R(\mathrm{d}\theta)$. From the properties of the Gamma process, the probability measure G a Dirichlet process, moreover G is independent of R. This suggests the following model, equivalent to the one above:

$$R \sim \Gamma(1, \alpha) \tag{9}$$

$$G(\mathrm{d}\theta) \sim \mathrm{DP}(\alpha) \tag{10}$$

$$E \sim \text{PoissProc}(R) \tag{11}$$

$$y_e \sim G \quad \forall e \in E \tag{12}$$

$$\mathbf{x}(t) = \sum_{e \in E} A y_e \delta_{(t-e)} \tag{13}$$

Our data is in a form that makes discrete-time modelling more natural, the Bernoulli approximation to the Poisson process suggests the following approximation: 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.

2 Inference

We perform inference in an online manner [5]. 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

$$z_t \sim Bern(p) \tag{14}$$

if $z_t == 1$

$$\gamma_t | \gamma_{1:t-1} \sim CRP \tag{15}$$

$$\theta_t | \gamma_t = i \sim \mathcal{N}(\theta_{t-1}, \Sigma) \tag{16}$$

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