# 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 $\Delta$, 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 number 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 neuron themselves emit continuous-time voltage traces, with the outputs of all neurons are 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 an appropriate kernel (the latter determines the shape of each action potential). We describe this in detail, starting first with the model for a single channel recording $X \equiv (x_1, \cdots, x_T)$.

We treat each neuron as stationary and memoryless, so that the set of spike times of each neuron is distributed as a homogeneous Poisson process. Comment on refractoriness or leave for discussion/future work. Similarly on the generalization to inhomogeneity? Let $r_i$ be the unkown firing rate for neuron $i$, and let $E_i$ be its spike times in increasing order; 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 involves a smooth perturbation in voltage about a resting state (without any loss of generality, we set the latter to zero). (figure? better biological description?). This perturbation forms the shape of the spike, and while this varies across neurons as well as across different spikes of the same neuron, each neuron has its own characteristic distribution over shapes. Figure? Comment on intrinsic change in waveform vs moving electrodes 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, offseting it by the time of the spike. 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 ( Figure? Comment on how this dictionary is obtained now, or in section on inference?). For the $i$th neuron, the $j$th 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 normally 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

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

(3)

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. Let $n(j)$ be the neuron to which the $j$th element of $E$ belongs, and let $p(j)$ index the position of the $j$th spike of $E$ in the spike train $E_{n(i)}$ of neuron $n(i)$. Then, we now have that we have

$$x(t) = \sum_{i=1}^{\infty} x_i(t) = \sum_{j=1}^{|E|} \sum_{k=1}^{K} y_{jk} A_k(t - e_j) \tag{4}$$

where

$$y_j \equiv \tilde{y}_{n(j)p(j)} \sim N(\mu_{n(j)}, \sigma_{n(j)}) \tag{5}$$

From the superposition property of the Poisson process [1], the overall spiking activity $E$ is a realization of a Poisson process with rate $R = \sum_{i=1}^{\infty} r_i$. The signal $x(t)$ is a functional of a marked Poisson process, where event $i$ is labelled by the neuron to which it is assigned ($n(i)$), and the shape of its spike waveform ($y_i$). From the properties of the Poisson process, it follows that the marks $n(i)$ are i.i.d. distributed with $P(n(i) = j) = \frac{r_j}{R}$. Given $n(i)$, $y_i$ is distributed as in equation 5.

As mentioned earlier, we let 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 have been well studied in the Bayesian nonparametrics community, where they form flexible and convenient priors over quantities like probability distributions, hazard functions etc. Besides being well understood theoretically, there is also a considerable literature on computational approaches for posterior inference.

Recall that each neuron is characterized by a pair $(r_i, \theta_i)$; 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 $\Theta$:

$$R(\mathrm{d}\theta) = \sum_{i=1}^{\infty} r_i \delta_{\theta_i} \tag{6}$$

For any subset $\Theta$ of $\Theta$, 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 (CRM). Such a random measure has the property that for any two disjoint subsets $\Theta_1$ 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 $\theta_i$'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'. The locations $\theta_i$ 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 ($\Gamma$P); this has Lévy intensity $r^{-1}\exp(-r\alpha)$. 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. The overall model is then:

$$R(\mathrm{d}\theta) \sim \Gamma\mathrm{P}(\alpha, H(\theta)) \tag{7}$$

$$E_i \sim \mathrm{PoissProc}(r_i) \quad i \text{ in } 1, 2, \cdots \tag{8}$$

$$\tilde{y}_{ij} \sim N(\mu_i, \Sigma_i) \quad i, j \text{ in } 1, 2, \cdots \tag{9}$$

$$x_i(t) = \sum_{j=1}^{|E_i|} \tilde{y}_{ij} A_j(t - e_{ij}) \tag{10}$$

$$X = \sum_{i=1}^{\infty} x_i \tag{11}$$

It will be more convenient from the point of inference to work with the marked Poisson process representation of equations 4 and 5. The overall 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 $\alpha$ respectively. The label $n(j)$ assigning events to neurons are drawn i.i.d. from a normalized Gamma process $G(\mathrm{d}\theta)$:

$$G(\mathrm{d}\theta) = \frac{r_j}{R} \tag{12}$$

$G(\mathrm{d}\theta$ is a random probability measure called a normalized random measure; for the Gamma process, this is a realization of the Dirichlet process. Moreover $G$ is independent of the total mass $R(\Theta)$, resulting in the following model equivalent to the one above:

$$R \sim \Gamma(1, \alpha) \tag{13}$$
$$G(\mathrm{d}\theta) \sim \mathrm{DP}(\alpha) \tag{14}$$
$$E \sim \mathrm{PoissProc}(R) \tag{15}$$
$$y_e \sim G \quad \forall e \in E \tag{16}$$
$$\mathbf{x}(t) = \sum_{e \in E} A y_e \delta_{(t-e)} \tag{17}$$

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

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

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