Bayesian Non Parametric Methods for Spike Detection in Calcium Imaging Data

Sara Capozio, Daniel Robin
Université Paris-Dauphine

Introduction

The paper proposed by Laura d'Angelo, Antonio Canale, Zhaoxia Yu and Michele Guindani, presents the application of nested Bayesian nonparametric models to measure neuronal response to different external stimuli in animals through the analysis of calcium signals. One of the techniques used in biometrics for analyzing neuronal activity is calcium imaging, a method used to record variations in calcium concentration within cells. This methodology involves the use of fluorescent probes that act as sensitive indicators to the presence of calcium. When neurons are activated and calcium floods into the cell generating concentration spikes, the fluorescent probes change their emission state producing a fluorescence trace. Analyzing this trace allows researchers to monitor neuronal activity in real-time and identify which neurons are active in response to which stimuli. It is important to emphasize that we are working with a highly imprecise variable. The fluorescence trace in calcium imaging, while valuable for monitoring neuronal activity, does not perfectly represent the trend of calcium concentration for several reasons. Firstly, the relationship between fluorescence and calcium is nonlinear, leading to a response that is not proportional to the variation in calcium concentration. Additionally, the complex dynamics of binding and release of fluorescent probes can cause delays in fluorescence response compared to actual calcium variations. Therefore, it is important to interpret this variable cautiously and consider its limitations and possible distortions. This analysis was conducted on a publicly available dataset from the Allen Brain Observatory on physiological activity in the mouse visual cortex in response to a series of visual stimuli. Each mouse is placed in front of a screen where different types of visual stimuli are shown, while neuronal activity is recorded. The aim of the study is to investigate how neurons at different depths in the visual areas respond to stimuli of different complexity through the application of a coherent nested hierarchical Bayesian finite mixture model. This model allows for the estimation of spike activity of each neuron and the reconstruction of spike distributions under various experimental conditions. In this report, we will review the main ideas of the proposed model, show you some initial results and explore an attempt of improvement of the model.

1 Priors and Hypothesis

Let us first explain how the model was built from a probabilistic perspective. The temporal behavior of the fluorescence trace is modeled through the definition of two main equations:

1) The observation and measurement equation:

$$y_t = b + c_t + \epsilon_t$$

where:
$-$ b: represents the baseline level of brain activity
$-$ c_t: the level of calcium concentration at time t
$-$ $\epsilon_t \sim N(0, \sigma^2)$ the noise

2) The transition equation that models the calcium level through a first-order auto-regressive model

$$c_t = \gamma c_{t-1} + A_t + \omega_t$$

where:
$-\gamma$ is the auto-regressive coefficient which controls the natural decay of calcium concentration. Assuming that the time series is stationary, we constrain this parameter to be in $]0;1[$
$-c_{t-1}$: calcium level at time $t-1$
$-A_t$:indicates the spike intensity at time t. When a spike occurs, the concentration increases with respect to the intensity of the neuronal activation.
$-\omega_t \sim N(0,\tau^2)$: is the noise in calcium concentration

Overall, the fluorescence trace follows a normal distribution:

$$y_t \sim N(b + \gamma c_{t-1} + A_t, \sigma^2 + \tau^2)$$

with the following fixed priors on the unknown parameters and with arbitrarily fixed hyper-parameters:

$-\frac{1}{\sigma^2} \sim \text{Gamma}(h_{1\sigma}, h_{2\sigma})$.
$-b \sim N(b_0, B_0)$
$-c_0 \sim N(0, C_0)$
$-\gamma \sim \text{Beta}(h_{1\gamma}, h_{2\gamma})$
$-\frac{1}{\tau^2} \sim \text{Gamma}(h_{1\tau}, h_{2\tau})$

It is known that neuronal activity can vary significantly in response to different experimental stimuli. The fundamental objective is therefore to create a hierarchical model that allows us to examine how neuronal activity is distributed across different scenarios. In practice, we want to group experimental data into clusters to identify if certain scenarios exhibit similar patterns of neuronal activity. The main objective of the analysis is therefore to infer the level of spike intensity A_t. Let J be the total number of settings considered, and let g_t be a discrete categorical variable taking values from $\{1,\dots,J\}$, indicating which scenario each recorded neuronal activity at time t belongs to. It is assumed that the spikes A_t follow stimulus-specific distributions. Each type of stimulus elicits a different neuronal response modeled by the distribution G_j: $A_t \,|\, g_t = j \sim G_j$
Despite the presence of various stimuli, some scenarios may exhibit similar patterns of neuronal activity. This means that over the J settings, we expect only K different clusters of neuronal behaviors (with $K < J$). The main objective is to cluster the probability distributions $G_1,\dots,G_j$ in order to identify similar patterns of neuronal activity across different experimental conditions. We model it as a general mixture of mixtures :

$$G_1,\dots,G_j \,|\, Q \sim Q, \quad Q = \sum_{k=1}^{K} \pi_k \delta_{G_{k^*}}$$

where:

$-\pi_1,\dots,\pi_K \sim \text{Dirichlet}_K\left(\frac{a}{K},\dots\frac{a}{K}\right)$,with $a > 0$

$-K$(number of clusters) $\sim$ Beta_negative_binomial

$-G_{1^*},\dots,G_{k^*}$ are the realizations of a finite mixture model.

For every behavioral cluster, the neuronal activity is characterized by different spikes intensities that may be shared by different settings. For instance there could be a level of "strong intensity neuronal activity" that is reached with probability q_1 in behavioral cluster 1 and with probability q_5 in behavioral cluster 5. To represent this aspect, we will define a set of spike intensities A_1^*, ..., A_L^* on which the behavioral clusters' distributions will take support. Mathematically,

$$G_k^* = \sum_{l=1}^{L} \omega_{l,k} \delta_{A_l^*}$$

where
$-(\omega_{1,k},\dots,\omega_{L,k}) \sim \text{Dirichlet}_L\left(\frac{\beta}{L},\dots,\frac{\beta}{L}\right)$, with $\beta > 0$.
$-A_l^* \sim G_0 = (1-p)\delta_0 + p \cdot \text{Gamma}(h_{A1}, h_{A2})$, a spike and slab mixture prior. We will carefully choose the parameters of the gamma distribution as we want to make sure we clearly separate the points with no activity from the spikes with an intensity adequately far from 0.
$-p$ (proportion of spikes) $\sim$ Beta(h_{1p}, h_{2p}), with $h_{1p} \ll h_{2p}$.

As a result we obtain a model where every setting is linked to a behavioral cluster of neuronal activity. Each behavioral cluster defines a distribution on neuronal activity spikes. Moreover these spikes of activity are shared by the clusters, thus representing different levels of activation a neuron can hit.

2 Posterior Inference

2.0.1 The method

Once the model has been defined, the authors suggest a method to find the posterior distribution of the parameters of the model. This interests us as we want to reproduce the whole underlying activity. The method mixes several statistical and Monte Carlo Markov Chains techniques such as Gibb's sampling, Kalman filtering and nested telescopic sampling. The entire sampling procedure used in the coding phase is reported below:

1. Kalman Filter: the calcium level c_t is sampled using a Kalman filter and a forward filtering backward sampling method. The Kalman filter provides a recursive estimate of the state variable based on the previous state estimate and the current observation, while taking into account the system dynamics and measurement noise.

2. The parameter b is sampled through Gibb's sampling step. The posterior distribution of the baseline level is given by $b \sim N(\frac{b_0}{B_0} + \frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t - c_t), \sqrt{\frac{1}{B_0} + \frac{T}{\sigma^2}})$

3. Similarly we use a single Gibb's sampling step for the parameters σ and τ. The posterior distributions for the two parameters are respectively

$\frac{1}{\sigma^2} \sim \text{Gamma}(h_{1\sigma} + \frac{T}{2}, h_{2\sigma} + \frac{1}{2}\sum_{t=1}^{T}(y_t - c_t - b)^2)$

$\frac{1}{\tau^2} \sim \text{Gamma}(h_{1\tau} + \frac{T}{2}, h_{2\tau} + \frac{1}{2}\sum_{t=1}^{T}(c_t - \gamma c_{t-1} - A_t)^2)$.

4. The parameter γ is sampled through a Metropolis-Hastings. At each iteration a new value of gamma is proposed and accepted with respect to the Metropolis-Hastings ratio.

5. The weight of the Dirichlet processes and the clusters allocating variables are sampled through a nested telescopic sampling.

6. Finally we use a Gibb's sampling step to sample a new parameter $p \sim \text{Beta}(h_{1p}+T-n_0, h_{2p}+n_0)$, where n_0 is the number of time stamps where $A_t = 0$.

At the end of this process, we have updated one time the whole model. After a few burn-in steps, we should start getting values approximately following the real posterior distribution of the model.

2.0.2 The results

The procedure described above was applied to the Allen Brain Observatory mouse brain data. The data represent the fluorescent trace detected by exposing a sample of mice to three different stimuli: static gratings, natural scene and natural movie. The number of scenario taken into account is J=4, with level 4 representing the category: 'absence of stimuli'. The clustering model reveals a similarity in the distribution of neural activity across the mouse exposed to natural scene and natural movies. However, for the category static grating, the distribution of spikes appeared to differ significantly. To analyze the spike variation across scenarios, a metric called firing rate was computed. This indicator measures how often the neurons activate during a specific visual stimulus calculating the number of detected spike per second. For the static grating stimulus the average posterior firing rate takes a value equal to 0.223, while the others two categories it takes a value equal to 0.419 and 0.511 respectively. In conclusion, it was demonstrated that there exists an association between neural activity and the complexity of the stimuli.

3 Extensions

3.0.1 Sensitivity analysis

The sensitivity analysis was conducted to assess how variations in the input values of the model affect the model's outcomes. Essentially, the main objective is to understand how sensitive the model's responses are to changes in its parameters or initial conditions. This approach aims to identify the most sensitive parameters and to define their range of variability. The hyper-parameters we focused on are those related to the definition of the spike and slab priors. One of the key points of the paper is to assign specific distributions to parameters A_t(intensity of the spikes) and p (proportion of spikes)

in order to enforce sparsity in the detection of spikes. The authors chose priors for these parameters capable of inducing a clear separation between baseline neuronal activity and neuronal response:
$A_l^* \sim G_0 = (1-p)\delta_0 + p \cdot \text{Gamma}(h_{A1}, h_{A2})$
p (proportion of spikes) $\sim \text{Beta}(h_{1p}, h_{2p})$

The choice of hyper-parameter values is crucial to enforce the sparsity, indeed the higher the shape parameter h_{A1} and the smaller h_{1p} is with respect to h_{2p}, the larger is the separation. During the analysis, the authors fixed the following values for the hyper-parameters: h_{A1}=8, h_{A2}=8, h_{1p}=1, h_{2p}=999. Through a simulation approach, we analyzed the robustness of the model by assessing how the posterior distribution changed as these parameters varied. Firstly, focusing on p, it was found that as the hyper-parameters changed, the posterior distribution for A_t remained very stable.

The inference made on this parameter appears to be highly robust to changes in the hyper-parameters. On the contrary, the inference made on the spike intensity values appears to be much more sensitive. Keeping the scale parameter constant, we analyzed the behavior of the posterior distribution of intensities as the shape parameter varied. After several simulations, it was noticed that the shape of the posterior distribution of A_t seemed to be highly variable when higher scale parameters (greater than 16) were chosen.

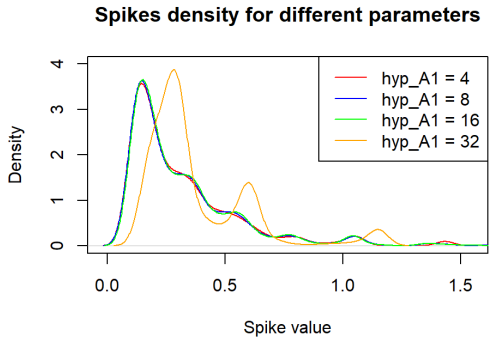

Figure 1: Param 1A

Since during the model validation phase, the empirical average intensities assumed a value of 1.5, it was decided to maintain scale and shape parameters consistent with the real-world results. For this reason, it is appropriate to consider shape parameter values between 8 and 16. Since in this range of variation, the specific choice of the hyper-parameter does not seem to have a significant impact on the posterior distribution, we can conclude that the inference made on A_t is robust.

3.0.2 Time dependency of the spikes

The authors applied the model without assuming a particular auto-correlation structure for the intensities A_t. This could represent an unrealistic scenario as typically spike intensities might exhibit an underlying auto-correlation structure that describes how the series evolves over time. For this reason a possible extension to this model is adding a time dependence on the values of the A_t. In order to check the robustness of our model against time-correlated spikes we tested its performances on a set of simulated data under the assumption that they follow an auto-regressive process. To model this auto correlation we did as follows. At every time t, if there was a spike at the previous time step, we increased the probability (by a fixed value rho) of having a nonzero spike. In this way, we have constructed a temporal dependency between spike levels that tends to exponentially decay as the time window increases. We assessed the performance of the model checking the value of the Rand Index in Distributional estimation (attributing each spike to the right cluster of distribution), and on Observational estimation (predicting the value of each spike intensity). As we can see by comparing the original scenario and the time-dependent one, the results are very close, with a decrease of performance and an increased variability in the time-dependent case. So the introduction of time-correlated data slightly weakens the model. Yet the model still achieves very good predictions.

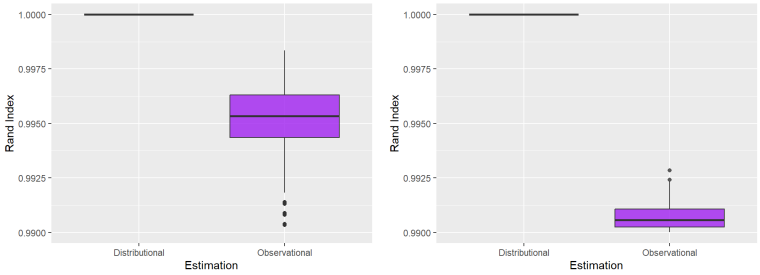

Figure 2: Rand Indexes of our simulations. On the left, the classical setting. On the right, simulation with time-dependent spike intensities

It is interesting to note that, as showed in the auto correlation plots, the model successfully catches the time dependence structure of the spike intensities, despite the absence of time correlated structure in its priors.

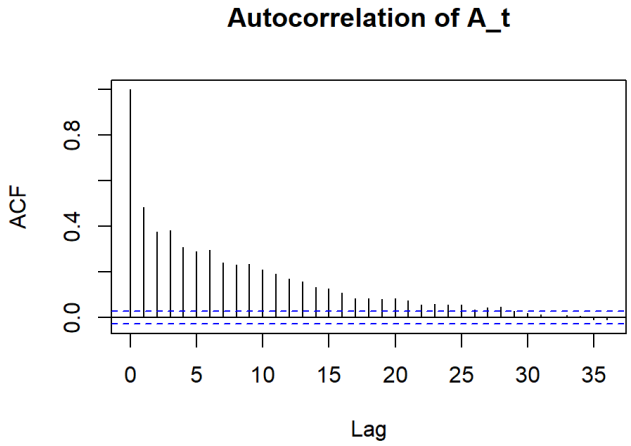

Figure 3: Autocorrelation plot of the posterior estimated spikes, for simulated data with time-dependency with rho=0.5

The main reason why the model appears to work under the assumption of auto-correlation among the data might be related to the fact that during the posterior inference, this property is preserved thanks to the strong information content of the sample. At the end, different scenarios belonging to the same cluster will exhibit the same distribution of spike intensities while sharing the same auto-correlation structure.

Conclusion

In conclusion, this report aimed to revisit the application of a hierarchical nested Bayesian model to reconstruct spike distributions under various experimental conditions from both theoretical and empirical perspectives. Through the application of sensitivity analysis, we assessed the robustness of the method with respect to the hyper-parameters used in the spike and slab priors, obtaining satisfying results. Then we explored a potential extension of the model, through the addition of time dependent auto-correlation of the spike intensities. We found that the model was robust to this type of structure in the empirical data as it was able to reproduce in the posterior inference the auto-correlation property of the spike intensities. Overall this study highlights the strength of this model in different situations, as no flaw was proven.