- Evaluator: Prof. Raffaele Argiento -

Main comments:

- "I think the Conclusions (or the Introduction) would benefit of some comment highlighting the specific advantages the Bayesian Nonparametric approach brings to the analysis of calcium imaging data. More specifically, which kind of conclusion/findings cannot be achieved with standard techniques? Which improvements the new approach brings to this applicative field."

  We thank the Evaluator for the suggestion: in the Conclusions we have stressed the role and importance of the Bayesian methodology to address the modeling challenges of the considered data.

- "The PhD candidate proposes a new algorithm for Poisson regression models that uses the Negative Binomial approximation. I'm wondering then, why one should uses a Poisson regression and not a Negative binomial regression to fit the responses? In other words, there is any specific reason to consider a Poisson regression for the Spike train data. Maybe a comparison between Poisson and Negative Binomial model will give strength the motivation for the proposed algorithm."

  We thank the Evaluator for raising this point. Despite we motivated the development of Chapter 2 with the analysis of spike train data, whose model goodness of fit would be appropriate to perform (e.g. to compare different model specification, such as negative binomial regression and ZIP, among others), the contribution has a methodological focus. The development of novel sampling schemes for Bayesian Poisson regression was motivated by the lack of efficient algorithms to perform posterior inference for this model. Hence this motivation goes beyond application to specific data sets or fields (as, for example, calcium imaging studies). Of course, in cases where a negative binomial distribution is more suited to describe the data, a different model, and thus a different sampling scheme should be used. However, in order to fit a Poisson regression (also within the scope of model comparison), our approach would represent a convenient algorithm to perform posterior inference.

- "Does the author think that the novel posterior sampling scheme for the Poisson regression could be extended to zero-inflated models."

  We thank the Evaluator for the interesting and stimulating question. Zero-inflated Poisson (ZIP) regression assumes that the distribution of the observed data can be expressed as a mixture of a Dirac mass at 0 and a Poisson distribution, i.e.,

$$f(y_i \mid \boldsymbol{\beta}, \pi_i) = \pi_i \delta_0(y_i) + (1 - \pi_i) e^{-e^{x_i^T \boldsymbol{\beta}}} e^{x_i^T \boldsymbol{\beta} y_i} \frac{1}{y_i!}.$$

  If π_i and $\boldsymbol{\beta}$ are assumed independent a priori, and if $\boldsymbol{\beta} \sim N_p(b, B)$, then the proposed algorithms could be applied without much additional effort. In particular, the proposed sampling strategy would be straightforward to apply if it is possible to implement a data augmentation scheme, where the latent allocation variables that indicate the component to which y_i is assigned are sampled.

If, for example, we consider the simplest setting where we assume $\pi_i = \pi \sim \text{Beta}(\alpha, \gamma)$ for all subjects, and we introduce the variables $v_i \in \{0, 1\}$, with $v_i = 1$ indicating the assignment to the degenerate component at zero, then

$$\Pr(v_i = 1 \mid y_i = 0, \pi, \boldsymbol{\beta}) = \frac{\pi}{\pi + (1 - \pi)\exp\{-e^{x_i^T \boldsymbol{\beta}}\}} \tag{1}$$
$$\Pr(v_i = 1 \mid y_i > 0, \pi, \boldsymbol{\beta}) = 0$$

and

$$f(y_i \mid v_i = 1) = \delta_0(y_i)$$
$$f(y_i \mid v_i = 0, \boldsymbol{\beta}) = \exp\{-e^{x_i^T \boldsymbol{\beta}}\} e^{x_i^T \boldsymbol{\beta} y_i} \frac{1}{y_i!}$$

Conditionally on $v_i = 0$, the setting is identical to that adopted in the Thesis.
In this setting, a Gibbs sampler algorithm is defined as:

1. Sample $\pi \mid v_1, \dots, v_n \sim \text{Beta}(\alpha + \sum_{i=1}^n v_i, \gamma + n - \sum_{i=1}^n v_i)$

2. For $i = 1, \dots, n$, sample $v_i \mid \pi, y_i, \boldsymbol{\beta}$, which is a Bernoulli with (conditional) probability given in Equation (1).

3. For the subgroup of observations with $v_i = 0$, sample the parameters $\boldsymbol{\beta}$ using one of the algorithms outlined in Chapter 2.
   This step can be performed without additional complexity because the update of $\boldsymbol{\beta}$ does not involve other variables (e.g. the Pólya-gamma data augmentation variables, which are integrated out). Thus, even if the set of y_i changes between iterations of the Gibbs sampler, the sampling of the $\boldsymbol{\beta}$ is not affected.

We have added a section where we discuss this extension of the model.

- " Section 1.3.1: I don't understand the need of introducing the component indicator S_{ik}. Indeed, in equation (1.2) the author introduces the cluster indicator c_i that is equivalent to $S_{ik}, k = 1, \dots K$. Moreover, it seems to me that in the rest of the work only the c_i variables are used."

  We thank the Evaluator for noticing this redundancy in the notation, the model specification was edited as suggested.

- "Equation (1.2) page 11. I suggest to write the first line of the model as follows $y_i | c_i, \theta_1^*, \dots, \theta_K^* \sim f(y_i | \theta_{c_i}^*)$."

  We changed as suggested.

- " '...we followed the unpublished report by Yee Whye Teh...' It is possible to refer better to this report. There are several books and papers introducing the Dirichlet Process. I would rather refer to an established books. Among the other I report here
  Müller, P., Quintana, F. A., Jara, A., & Hanson, T. (2015). Bayesian nonparametric data analysis. New York: Springer."

  We thank the Evaluator for the suggestion, and we updated the references.

- "Section 1.3.2: In this section the author refers to the number of clusters using the notation K. This is a bit misleading because in the Section 1.3.1 and in Section 1.3.3 K denotes the number of components of the mixture. I suggest to denote by K^+ the number of clusters also in this Section."

  We thank the Evaluator for the useful indication, we corrected the notation accordingly.

- "Finite mixture with random number of components are a very old and celebrated class of

models. Arguably, two of the seminal papers that strongly have contributed to the popularity of this class are

Richardson, S., & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). Journal of the Royal Statistical Society: series B (statistical methodology), 59(4), 731-792.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. Annals of statistics, 40-74.

It is very unclear to me why in many recent works (e.g. in this Thesis) researchers refer to the class of mixtures models with random number of components as mixture of finite mixture. I think the latter name is a misleading reference that do not recognize the early and seminal works on the topic."

We have added the suggested references, which certainly contribute to provide a comprehensive review of these models. We agree that seminal works such as those just mentioned have been fundamental for research concerning mixtures with a random number of components. However, in this thesis we have exploited the recent literature (in particular, Miller and Harrison, 2018; and Frühwirth-Schnatter et al., 2021), where this class of model is referred to using the name of "mixture of finite mixtures". Hence, in line with these works, we have decided to use the same definition.

- " 'Considering fixed parameters as in Miller and Harrison (2018), where a $\text{Dirichlet}_K(1, \ldots, 1)$ is used regardless of the value of K, leads to a prior expected number of clusters $E(K^+)$ close to $E(K)$ for many priors p(K).'
  I really disagree with this sentence. Can the author provide some illustrative example? Moreover, I think that the $\text{Dirichlet}_K(\gamma, \ldots, \gamma)$ prior, with γ independent from M, can yields to very flexible prior on K^+. See for instance, Argiento & De Iorio (2019), or the AntMAN package (https://cran.r- project.org/web/packages/AntMAN/index.html)"

  We agree with the Evaluator, that sentence was not very clear and it could lead to some confusion; we have clarified it to avoid misunderstandings. Illustrative examples can be found in the paper by Frühwirth-Schnatter et al. (2021).

- "Section 2.1.3: Is the adaptive importance sampling scheme discussed in this Section related with the sequential importance sampling.
  Del Moral, P., A. Doucet, and A. Jasra (2006). Sequential Monte Carlo samplers. J. R. Stat. Soc. Series B Stat. Methodol. 68(3), 411–436.
  I think a sentence to clarify connections (or absence of connections) should be added here."

  We thank the Evaluator for the useful and interesting reference, we have discussed connection with this method in Section 2.1.3. From my understanding, in the work of Del Moral, P., A. Doucet, and A. Jasra (2006) the main goal is to define a sampling scheme for sequences of distributions π_n, for n varying in $\mathbb{T} = \{1, \ldots, p\}$. Their work hence focuses on updates of the sampling procedure to exploit knowledge about the target density π_{n-1} to sample from π_n in an efficient way. Instead, in the proposed algorithms we use information on the previous state of the chain to inform the proposal density. Although there are some analogies between their approach and our proposed algorithm, as they both try to maximize the available information to improve the sampling scheme, the main objective is different.

- "Section 3.3 (more precisely pg 39). 'The result attained by the proposed fCAM". It is not clear to me how spike's detection is performed for the simulation study. In the subsequent Section 3.4 (pg 42), the author states that a spike is identified if posterior probability of a spike at time t is greater than $\kappa = 75.5\%$.' I'm wondering if the same threshold is used also in the simulation study. If this is the case, how sensitive are the results to the choice of this

threshold?"

We thank the Evaluator for this question. We have clarified the spike detection procedure used in the simulation study by moving the discussion about the false discovery rate in that section. Moreover, we have added an analysis of the sensitivity of spike detection to the choice of the threshold (see page 41).

- "I like Chapter 4, but, as the authors also mentioned in the conclusions I think that the proposed approach can be strongly improved from a computational point of view. I think the PhD candidate could elaborate a bit more about this issue in this chapter."

We agree that the discussion about the computational aspects was somehow approximate. At the beginning of Section 4.3, we have discussed more in detail the computational issues of the current approach, and we have outlined some possible strategies to improve it.

- Evaluator: Prof. Peter Müller -

The Evaluator did not request any edit of the Thesis.