Response to the evaluators' comments and suggestions for the thesis
"Bayesian modeling of calcium imaging data"
by
Laura D'Angelo

- 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."

  Response of the PhD candidate.

- "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."

  Response of the PhD candidate.

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

  Response of the PhD candidate.

- " 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 $S_{ik}, k = 1, ...K$. 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- " '...we followed the unpublished report by Yee Whye Teh...'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."

  Response of the PhD candidate.

- " '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)"

  Response of the PhD candidate.

- "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."

  Response of the PhD candidate.

- "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 thank the Evaluator for the valuable suggestion. 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 improvements.

- Evaluator: Prof. Peter Müller -

The Evaluator did not request any edit of the Thesis.