Valutazione/Review Id: **684660**

1 VALUTAZIONE/REVIEW

1.1 Dati del candidato per cui si sta eseguendo la valutazione / PhD student Personal Data

Cognome /
Surname

D'ANGELO

Nome /
Name

LAURA

Titolo della tesi / Title of the
Thesis

Modellazione bayesiana di dati di calcium imaging

Cognome supervisore / Supervisor
surname

CANALE

Nome supervisore / Supervisor
name

ANTONIO

1.2 Dati del valutatore / Evaluator's data

Nome/Name

Raffaele

Cognome/Surname

ARGIENTO

E-mail

raffaele.argiento@unibg.it

Ente di appartenenza/Evaluator's
institution

Università degli studi di Bergamo

1.3 Valutazione / Review

Il titolo è coerente con il contenuto dell'elaborato?
Does the title adequately convey the content of the dissertation?

Si / Yes

Valore scientifico complessivo (originalità, rilevanza dello specifico ambito di ricerca, accuratezza)
Overall scientific merit: originality, relevance in the specific field, completeness

Eccellente / Excellent

Introduzione e bibliografia (l'introduzione fornisce informazioni sufficienti sull'argomento di ricerca della tesi, i riferimenti bibliografici citati sono appropriati e completi)
Introduction and bibliography: the introduction contains enough information on the object of the thesis, bibliographical references are suitable and complete

Eccellente / Excellent

Metodologia - Analisi dei dati (i metodi sono appropriati, l'analisi dei dati è adeguata ed esaustiva - se pertinente)
Methodology - Data Analysis: appropriate methods / suitable and exhaustive data analysis (if applicable)

Eccellente / Excellent

Descrizione del lavoro di ricerca
Research work description

Eccellente / Excellent

Risultati (i risultati sono presentati in modo chiaro e convincente, il numero e la qualità dei grafici e delle immagini sono adeguati)
Results: convincing and clearly presented results / suitable number and quality of tables and figures

Eccellente / Excellent

Discussione e conclusioni (interpretazione corretta dei risultati, contributo all'avanzamento nel campo della ricerca)
Discussion and Conclusions: correct interpretation of the results / advance of the results in the research field

Eccellente / Excellent

Valutazione finale
Overall evaluation

Il candidato può essere ammesso all'esame finale ma la tesi richiede minime correzioni/integrazioni; non è necessaria un'ulteriore valutazione da parte del valutatore.
The candidate can be admitted to the final examination but the thesis requires minor revisions; no addition review is required by the evaluator

Commenti e suggerimenti - Esplicitare con chiarezza le modifiche/integrazioni richieste
Specific comments and suggestions - Clarify the corrections/additions requested

I would like to congratulate Laura for the excellent work. I read her thesis with care and i think it is well written and interesting.

I have two main comments:

a) 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.

b) 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.

Other points are listed below:

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

ii) Section 1.3.1
I don't understand the need of introducing the component indicator Sik. 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.

iii) Equation (1.2) page 11.
I suggest to write the first line of the model as follows
$y_i|c_i,\theta*_1,\dots,\theta*_k \sim f(y_i|\theta^*_{c_i}$

iv) "...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.

v) 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.

vi) 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.

vii) pg 17
"Considering fixed parameters as in Miller and Harrison (2018), where
a Dirichlet$_K(1,\dots, 1)$ 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 Dirichlet$_K(\gamma,\dots, \gamma)$ 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)

viii) 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.

ix) 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?

x) 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.

PHD FINAL EXAMINATION 2022 - REVIEW

Valutazione/Review Id: **686222**

1 VALUTAZIONE/REVIEW

1.1 Dati del candidato per cui si sta eseguendo la valutazione / PhD student Personal Data

Cognome / Surname

D'ANGELO

Nome / Name

LAURA

Titolo della tesi / Title of the Thesis

Modellazione bayesiana di dati di calcium imaging

Cognome supervisore / Supervisor surname

CANALE

Nome supervisore / Supervisor name

ANTONIO

1.2 Dati del valutatore / Evaluator's data

Nome/Name

Peter

Cognome/Surname

Mueller

E-mail

pmueller@math.utexas.edu

Ente di appartenenza/Evaluator's institution

Univ Texas at Austin

1.3 Valutazione / Review

Il titolo è coerente con il contenuto dell'elaborato?
Does the title adequately convey the content of the dissertation?

Si / Yes

Valore scientifico complessivo (originalità, rilevanza dello specifico ambito di ricerca, accuratezza)
Overall scientific merit: originality, relevance in the specific field, completeness

Molto buono / Very good

Introduzione e bibliografia (l'introduzione fornisce informazioni sufficienti sull'argomento di ricerca della tesi, i riferimenti bibliografici citati sono appropriati e completi)
Introduction and bibliography: the introduction contains enough information on the object of the thesis, bibliographical references are suitable and complete

Eccellente / Excellent

Metodologia - Analisi dei dati (i metodi sono appropriati, l'analisi dei dati è adeguata ed esaustiva - se pertinente)
Methodology - Data Analysis: appropriate methods / suitable and exhaustive data analysis (if applicable)

Eccellente / Excellent

Descrizione del lavoro di ricerca
Research work description

Eccellente / Excellent

Risultati (i risultati sono presentati in modo chiaro e convincente, il numero e la qualità dei grafici e delle immagini sono adeguati)
Results: convincing and clearly presented results / suitable number and quality of tables and figures

Molto buono / Very good

Discussione e conclusioni (interpretazione corretta dei risultati, contributo all'avanzamento nel campo della ricerca)
Discussion and Conclusions: correct interpretation of the results / advance of the results in the research field

Eccellente / Excellent

Valutazione finale
Overall evaluation

Il candidato può essere ammesso all'esame finale
The candidate can be admitted to the final examination

Commenti e suggerimenti - Esplicitare con chiarezza le modifiche/integrazioni richieste

Specific comments and suggestions - Clarify the corrections/additions requested

Very nice work. The overall theme is fully model-based inference for fluorescence trace data for neuron activity. The proposed models move the field forward by going beyond currently used models in several ways, including a novel hierarchical mixture of finite mixtures for spike train patterns under different experimental conditions (Chapter 3) and a novel framework for clustering underlying longitudinal binary data with a spatial regression.
All models require highly computation-intensive implementation of inference. For example, posterior inference for spike train data under a GLM model with Poisson sampling becomes challenging. A novel approach to posterior simulation in such problems is introduced that is based on the Poly-gamma model augmentation of (negative) binomial models.
Overall, very nice work, with clear scope for developing a rich future related research program.

There is really not much to be critiqued. The work includes important new contributions in several areas of current statistics research, including computational statistics (the Poly-gamma data augmentation), nonparametric Bayesian model construction (the nested gMFM model) and model-based spatial clustering. All three major contributions develop new ideas with good scope for future research.
The nature of this work is close to my interests, and the style is close to my tastes in statistics research. In particular, I appreciate the focus on highly structured stochastic models to address important scientific research questions. The effort is clearly targeting the solution of important scientific research questions, and addressing the challenges that arise from implementing inference in these models. It is then natural that this approach does not lead to problems that are simple and sufficiently stylized to prove interesting mathematical results. Overall, the candidate has credibly shown an ability for scientific and scholarly work, excellent knowledge of the literature, familiarity with research methods including in particular computation and simulation, and importantly, good judgment in identifying promising focused research questions and gaps in the current literature.