



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
Corso di Dottorato di Ricerca in Scienze Statistiche
Ciclo XXXIV

Thesis Title

Coordinatore del Corso: Prof. Nicola Sartori

Supervisore: Prof. Antonio Canale

Co-supervisore: Prof. Michele Guindani

Dottoranda: Laura D'Angelo

Day Month Year of submission

Abstract

Abstract content.

Sommario

Contenuto del sommario.

Contents

List of Figures	vii
List of Tables	viii
Introduction	1
Overview	1
Main contributions of the thesis	2
1 Background: statistical modeling of calcium imaging data	3
1.1 Overview of calcium imaging data	3
1.1.1 Deconvolution methods	4
1.2 Data sets	5
1.2.1 Allen Brain Observatory data	5
1.2.2 Altri dati?	6
1.3 A brief review of some Bayesian nonparametric models	6
1.3.1 Finite mixture models	6
1.3.2 Dirichlet process mixtures	7
1.3.3 Mixtures of finite mixtures	9
1.3.4 Bayesian nonparametric models for nested data	9
2 Efficient posterior sampling for Bayesian log-linear models	11
2.1 Algorithm	11
3 Bayesian nonparametric analysis for partially exchangeable single-neuron data	13
3.1 Model specification	13
3.2 Posterior computation	13
Bibliography	15

List of Figures

List of Tables

Introduction

Overview

A fundamental but unsolved problem in neuroscience is understanding the functioning of neurons and neuronal networks in processing sensory information, generating locomotion, and mediating learning and memory. The investigation of the structure and function of the nervous system can be dated back to the nineteenth century with the invention of the technique of silver impregnation by Camillo Golgi in 1873, which allowed the visualization of individual neurons (Drouin *et al.*, 2015). The technique initiated the study of the microscopic anatomy of the nervous system, and the investigation of how neurons organize to form the brain. Ever since there has been a significant research effort both to discover the cellular properties of the nervous system, and to characterize behaviors and correlate them with activity imaged in different regions of the brain. However, many scientists recognize that despite the innovative techniques developed to observe and analyze neurons, we are still facing an “explanatory gap” between the understanding of elemental components and the outputs that they produce (Parker, 2006, 2010; Dudai, 2004). That is, we know a lot about the components of the nervous system, but still we have little insight into how these components work together to enable us to think, remember, or behave. One of the reasons of this gap is the availability of a huge quantity of data, but a lack of tools to integrate these data in order to obtain a coherent picture of the brain functioning (Parker, 2010).

The technological developments of the last few decades have opened fundamentally new opportunities to investigate the nervous system. Large neuronal networks can now be visualized using *in vivo* high-resolution imaging techniques, which permit to record the neuronal activity in freely moving animals over long periods of time. In this thesis, we focus on data resulting from the application of the two-photon calcium imaging technique. Calcium ions generate intracellular signals that determine a large variety of functions in all neurons: when a neuron fires, calcium floods the cell and produces a transient spike in its concentration (Grienberger and Konnerth, 2012). By using genetically encoded calcium indicators, which are fluorescent molecules that react when binding to the calcium ions, it is possible to optically measure the level of calcium by analyzing the observed fluorescence trace. However,

extracting these fluorescent calcium traces is just the first step towards the understanding of brain circuits: how to relate the observed pattern of neuronal activity with its output remains an open problem of research.

Main contributions of the thesis

Chapter 1

Background: statistical modeling of calcium imaging data

1.1 Overview of calcium imaging data

Calcium ions generate intracellular signals that control key functions in all types of neurons. At rest, most neurons have an intracellular calcium concentration of about 100 nM; however, during electrical activity, the concentration can rise transiently up to levels around 1000 nM (Berridge *et al.*, 2000). The development of techniques that enable the visualization and quantitative estimation of the intracellular calcium signals have thus greatly enhanced the investigation of neuronal functioning. The development of calcium imaging techniques involved two parallel processes: the development of calcium indicators, which are fluorescent molecules that react when binding to the calcium ions, and the implementation of the appropriate imaging instrumentation, in particular, the introduction of two-photon microscopy (Denk *et al.*, 1990). In recent years, the innovation achieved in these two fields has allowed for real-time observation of biological processes at the single-cell level simultaneously for large groups of neurons (Grienberger and Konnerth, 2012).

The output two-photon calcium imaging is a movie of time-varying fluorescence intensities, and a first complex pre-processing phase deals with the identification of the spatial location of each neuron in the optical field and source extraction (Mukamel *et al.*, 2009; Dombek *et al.*, 2010). The resulting processed data consist of a fluorescent calcium trace for each observable neuron in the targeted area which, however, is only a proxy of the underlying neuronal activity. Hence further analyses are needed to deconvolve the fluorescence trace to extract the spike train (i.e. the series of recorded firing times), and to try to explain how these firing events are linked with the experiment that generated that particular pattern of activity.

1.1.1 Deconvolution methods

There is currently a rich literature of methods addressing the issue of deconvolving the raw fluorescent trace to extract the spike train. A successful approach is to assume a biophysical model to relate the spiking activity to the calcium dynamics, and to the observed fluorescence. Vogelstein *et al.* (2010) proposed a simple but effective model that has later been adopted by several authors (Pnevmatikakis *et al.*, 2016; Friedrich and Paninski, 2016; Friedrich *et al.*, 2017; Jewell and Witten, 2018; Jewell *et al.*, 2019). The model considers the observed fluorescence as a linear (and noisy) function of the intracellular calcium concentration; the calcium dynamics is then modeled using an autoregressive process with jumps in correspondence of the neuron's firing events. Denoting with y_t the observed fluorescence trace of a neuron and with c_t the underlying calcium concentration, for time $t = 1, \dots, T$, the model can be written as

$$\begin{aligned} y_t &= b + c_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2), \\ c_t &= \gamma c_{t-1} + A_t + w_t, \quad w_t \sim \mathcal{N}(0, \tau^2), \end{aligned} \tag{1.1}$$

where b models the baseline level of the observed trace and ϵ_t is a Gaussian measurement error. In the absence of neuronal activity, the true calcium concentration c_t is considered to be centered around zero. The parameter A_t captures the neuronal activity: in the absence of a spike ($A_t = 0$), the calcium level follows a AR(1) process controlled by the parameter γ ; when a spike occurs, the concentration increases instantaneously of a value $A_t > 0$. A challenge remains estimating the neuronal activity A_t in a precise and computationally efficient way.

Vogelstein *et al.* (2010) assume that all spikes have a fixed amplitude, and interpret the parameter A_t as the *number* of spikes at time t . Following this definition, they place a Poisson prior distribution on A_t ; however, the maximum a posteriori estimation of the spike train using a Poisson distribution is computationally intractable. Hence they search an approximate solution by replacing the Poisson distribution with an exponential distribution of the same mean. This leads to some loss of interpretation of the parameters A_t , as now they are no longer integer values but rather non-negative real numbers, but turns the problem into a convex optimization, which can be solved efficiently. Adopting this approach leads to solving a non-negative lasso problem for estimating the calcium concentration, where the L_1 penalty enforces sparsity of the neural activity. Efficient algorithms to obtain a solution of this problem were also proposed by Pnevmatikakis *et al.* (2016), Friedrich and Paninski (2016), and Friedrich *et al.* (2017).

A different perspective is instead proposed by Jewell and Witten (2018) and Jewell *et al.* (2019): rather than interpreting A_t in model (1.1) as the number of spikes at the t -th timestep, they interpret its sign as an indicator for whether or not *at least one* spike occurred, that is, $A_t = 0$ indicates no spikes at time t , and $A_t > 0$ indicates the occurrence of at least one spike. The model so formulated includes an indicator variable, which corresponds to using an L_0 penalization and which makes the optimization problem highly non-convex. In their work,

Jewell and Witten (2018) and Jewell *et al.* (2019) develop fast algorithms to compute the spike trains under these assumptions. Jewell and Witten (2018) assert that the solutions discussed by Vogelstein *et al.* (2010), Friedrich and Paninski (2016), and Friedrich *et al.* (2017) can actually be seen as convex relaxations of this optimization problem, to overcome the computational intractability of the L_0 penalization.

Finally, Pnevmatikakis *et al.* (2013) propose a fully Bayesian approach. Although less computationally efficient than optimization methods, it allows to obtain a posterior distribution of all model parameters instead of just a point estimate, hence improving uncertainty quantification. Differently from previous models, they define the parameter A_t as the *amplitude* of a spike at time t , taking values in the non-negative real numbers. They formulate the presence/absence of a spike and its amplitude by using the product of a Bernoulli random variable (taking value 0 if there is no spike at time t , and 1 otherwise) with a half-Gaussian random variable (modeling the positive amplitudes). However, they do not explicitly assume sparsity of the spikes.

1.2 Data sets

Una frase introduttiva? Parlo dei dati dell'Allen Brain Observatory e poi ci sarebbe da mettere i nuovi dati se riesco a fare qualcosa del progetto 3...

1.2.1 Allen Brain Observatory data

The Allen Brain Observatory (Allen Institute for Brain Science, 2016) is a public large data repository for investigating how sensory stimulation is represented by neural activity in the mouse visual cortex in both single cells and populations. The project aims to provide a standardized and systematic survey to measure and analyze visual responses from neurons across cortical areas and layers, utilizing transgenic Cre lines to drive expression of genetically encoded fluorescent calcium indicators, and measured by *in vivo* two-photon calcium imaging.

The study is an extended survey of physiological activity in the mouse visual cortex in response to a range of visual stimuli (Allen Brain Observatory, 2017). Each mouse is placed in front of a screen where different types of visual stimuli are shown, while the mouse's neuronal activity is recorded. The stimuli vary from simple synthetic images such as locally sparse noise or static gratings, to complex natural scenes and movies. The goal of the study is to investigate how neurons at different depths in the visual areas respond to stimuli of different complexity. Specifically, each neuron in the visual cortex can be characterized by their *receptive field*, i.e. the features of the visual stimulus that trigger the signaling of that neuron. Hence, it is of critical interest to devise methods that allow inferring how the neuronal response varies under the different types of visual stimuli.

1.2.2 Altri dati?

Paragrafo qui.

1.3 A brief review of some Bayesian nonparametric models

In this section we review some statistical tools that will be employed in this thesis in the analysis of calcium imaging data. The purpose of this section is not to provide a comprehensive review, but rather to outline the theoretical framework we adopted and fix some notation. The core topic will be the Bayesian methodology, with a focus on Bayesian nonparametric models.

1.3.1 Finite mixture models

We start our discussion by reviewing finite mixtures. Although they are not part of the Bayesian nonparametric methodology, they provide the starting point for many models that we will review in the following. The content of this brief overview on finite mixtures is largely based on the dedicated chapter in Gelman *et al.* (2013).

Mixtures are a popular tool to model heterogeneous data, characterized by the presence of subpopulations within the overall population. In many practical problems the data are collected under different conditions – unfortunately, it is not always possible to have information on the subpopulation to which each individual observation belongs. Mixture models can be used in problems of this type, where the population consists of a number of latent subpopulations, within each of which a relatively simple model can be applied.

Denote the observed data as a vector of n units $\mathbf{y} = (y_1, \dots, y_n)$; also, assume that the n observations are exchangeable, meaning that the joint probability density $p(y_1, \dots, y_n)$ is invariant to permutations of the indices. In the framework of finite mixtures, we assume that the population is made of $K \leq n$ subpopulations, with K known and fixed. We assume that within each of these groups, the distribution of $y_i, i = 1, \dots, n$, can be modeled as $f(y_i | \theta_k^*)$, for $k = 1, \dots, K$. Usually a common parametric family is assumed for all these component distributions, which however depend on specific parameter vectors θ_k^* . The last missing piece to construct a mixture model is the parameter describing the proportion of population from each component k : we denote this parameter with π_k , satisfying $\sum_{k=1}^K \pi_k = 1$. Denoting the full vectors of parameters as $\boldsymbol{\theta}^* = (\theta_1^*, \dots, \theta_K^*)$ and $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$, the data distribution for observation i can be formulated as

$$p(y_i | \boldsymbol{\theta}^*, \boldsymbol{\pi}) = \pi_1 f(y_i | \theta_1^*) + \dots + \pi_K f(y_i | \theta_K^*). \quad (1.2)$$

In mixture models it is convenient to think of the component indicators as missing data, and to impute them to obtain a much simpler form of the data distribution. Hence we introduce

the indicator S_{ik} of component k for observation i , with

$$S_{ik} = \begin{cases} 1 & \text{if } y_i \text{ is drawn from component } k \\ 0 & \text{otherwise.} \end{cases}$$

Given π , the distribution of $\mathbf{S}_i = (S_{i1}, \dots, S_{iK})$ is Multinomial($1; \pi_1, \dots, \pi_K$). Conditionally on \mathbf{S}_i , the data distribution of y_i is simply $p(y_i | \mathbf{S}_i, \theta^*) = \prod_{k=1}^K f(y_i | \theta_k^*)^{S_{ik}}$; moreover, given $\mathbf{S} = (\mathbf{S}_1, \dots, \mathbf{S}_n)$, the y_i are assumed to be independent. The joint density of the observed data and the unobserved indicators, conditionally on the model parameters, can now be written as

$$p(\mathbf{y}, \mathbf{S} | \theta^*, \pi) = p(\mathbf{y} | \mathbf{S}, \theta^*) p(\mathbf{S} | \pi) = \prod_{i=1}^n \prod_{k=1}^K \{\pi_k f(y_i | \theta_k^*)\}^{S_{ik}}.$$

Having defined the data distribution, we need to specify adequate prior distributions on the model parameters π and θ^* . The prior G_0 on θ^* is usually chosen depending on the specific application and on the basis of the component distribution f . For the mixture proportions π_k , the conjugate and most natural prior distribution is the Dirichlet distribution, $\pi \sim \text{Dir}_K(\alpha_1, \dots, \alpha_K)$.

It is possible to rewrite equation (1.2) in a hierarchical way by thinking that each observation y_i is associated with a parameter θ_i , where these parameters are drawn from a discrete distribution G with support on the K locations $\{\theta_1^*, \dots, \theta_K^*\}$. The model then becomes, for $i = 1, \dots, n$

$$\begin{aligned} y_i | \theta_i &\sim f(y_i | \theta_i) \\ \theta_i | \theta^*, \pi &\sim G = \sum_{k=1}^K \pi_k \delta_{\theta_k^*} \\ (\pi_1, \dots, \pi_K) &\sim \text{Dir}_K(\alpha_1, \dots, \alpha_K) \\ (\theta_1^*, \dots, \theta_K^*) &\sim G_0. \end{aligned} \tag{1.3}$$

1.3.2 Dirichlet process mixtures

Nonparametric mixtures extend model (1.3) by placing a nonparametric prior on G . The most common prior on random probability measures is the Dirichlet process (DP), introduced by Ferguson (1973, 1974). Draws from a DP are discrete distributions with probability one, hence they turned out useful as flexible mixing measures in discrete mixtures.

A random distribution G on Θ is said to follow a DP prior with base measure G_0 and concentration parameter α , denoted $G \sim \text{DP}(\alpha, G_0)$, if for any partition $\{B_1, \dots, B_H\}$ of Θ ,

$$(G(B_1), \dots, G(B_H)) \sim \text{Dir}_H(\alpha G_0(B_1), \dots, \alpha G_0(B_H)).$$

The success of the DP mainly arises from two appealing characteristics: its large support, with respect to the space of probability distributions, and tractability of the posterior distribution.

Closely related to this last aspect is the conjugacy property of the DP: as the finite dimensional Dirichlet distribution is conjugate to the multinomial likelihood, the DP is conjugate with respect to i.i.d. sampling, that is, with respect to a completely unknown distribution from i.i.d. data. More precisely, if we take $\{\theta_1, \dots, \theta_n\}$ a sequence of independent draws from $G \sim \text{DP}(\alpha, G_0)$, then the posterior distribution of G given these observed values is still a DP. Letting again $\{B_1, \dots, B_H\}$ be a finite measurable partition of Θ , and letting n_h be the number of observed values in B_h , for $h = 1, \dots, H$, the posterior distribution is given by

$$(G(B_1), \dots, G(B_H)) \mid \theta_1, \dots, \theta_n \sim \text{Dir}_H(\alpha G_0(B_1) + n_1, \dots, \alpha G_0(B_H) + n_H).$$

In other terms, the posterior distribution is still DP with updated parameters:

$$G \mid \theta_1, \dots, \theta_n \sim \text{DP}\left(\alpha + n, \frac{\alpha G_0 + \sum_{i=1}^n \delta_{\theta_i}}{\alpha + n}\right).$$

The DP admits several nice representations. An intuitive constructive definition of a DP random probability measure is given by Sethuraman (1994) and is based on the discrete nature of the process, which can be represented as a weighted sum of point masses. This definition states that if $G \sim \text{DP}(\alpha, G_0)$, then it can be expressed as follows:

$$\begin{aligned} \beta_k &\sim \text{Beta}(1, \alpha), & \pi_k &= \beta_k \prod_{l=1}^{k-1} (1 - \beta_l) \\ \theta_k^* &\sim G_0, & G &= \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k^*}. \end{aligned}$$

From this representation it is straightforward to extend the finite mixture framework described above to nonparametric mixtures: following the structure of equation (1.3), DP mixtures (DPM) can be written as

$$\begin{aligned} y_i \mid \theta_i &\sim f(y_i \mid \theta_i) \\ \theta_i &\mid G \sim G \\ G &\sim \text{DP}(\alpha, G_0). \end{aligned}$$

Differently from the mixtures described in the previous section, DPM are an infinite mixture model, as they assume a countably infinite number of components. However, because the π_k 's decrease exponentially quickly, only a small number of components will be used to model the data a priori: in the following, we will define *clusters* these non-empty components. In general, the prior expected number of clusters, which corresponds to the number of distinct θ_i in a sample of n values, is approximately equal to $\alpha \log(1 + n/\alpha)$. This means that while in finite mixture models the number of clusters has to be fixed in advance, in DP mixture model,

the number of clusters is determined by the data and can be inferred.

1.3.3 Mixtures of finite mixtures

To avoid fixing the number of clusters, a different approach to DP mixtures is to consider finite mixtures with a prior on the number of components. Although this may seem as the most natural way to infer the number of groups, application of mixtures of finite mixtures (MFM) has long been hindered by the difficulty of performing posterior inference. Several inference methods have been proposed for this type of model (McCullagh and Yang, 2008; Nobile, 2004; Nobile and Fearnside, 2007; Richardson and Green, 1997) often exploiting the reversible jump Markov chain Monte Carlo technique.

More recently, different models have been proposed inspired by nonparametric mixtures. Miller and Harrison (2018) explicitly derive properties of MFM analogous to those exhibited by DPM. They consider a finite mixture model similar to the one defined in equation (1.3), with a prior $p(K)$ on the number of clusters. The limit of their approach is that it is assumed a fixed parameter α for the Dirichlet distribution, regardless of the dimension K .

Another approach is discussed by Malsiner-Walli *et al.* (2016) and Frühwirth-Schnatter and Malsiner-Walli (2019), based on the use of sparse mixtures. Similarly to DPM, in this formulation they distinguish between the mixture components and the clusters, which are defined as the components actually used by the data. The number of components K is fixed to a large and clearly overfitting value, and a Dirichlet prior with small parameter on the mixture weights ensures that the (random) number of clusters K_+ will take values smaller than K with high probability a priori and also a posteriori.

1.3.4 Bayesian nonparametric models for nested data

Chapter 2

Efficient posterior sampling for Bayesian log-linear models

2.1 Algorithm

Chapter 3

Bayesian nonparametric analysis for partially exchangeable single-neuron data

3.1 Model specification

3.2 Posterior computation

Bibliography

- Allen Brain Observatory (2017) Technical whitepaper: stimulus set and response analyses.
- Allen Institute for Brain Science (2016) Allen brain observatory. Available at: <http://observatory.brain-map.org/visualcoding>.
- Berridge, M. J., Lipp, P. and Bootman, M. D. (2000) The versatility and universality of calcium signalling. *Nature Reviews Molecular Cell Biology* **1**, 11 – 21.
- Denk, W., Strickler, J. H. and Webb, W. W. (1990) Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76.
- Dombeck, D. A., Harvey, C. D., Tian, L., Looger, L. L. and Tank, D. W. (2010) Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. *Nature Neuroscience* **13**, 1433 – 1440.
- Drouin, E., Piloquet, P. and Péréon, Y. (2015) The first illustrations of neurons by Camillo Golgi. *The Lancet Neurology* **14**(6), 567.
- Dudai, Y. (2004) The neurosciences: the danger that we will think that we have understood it all. In *The new brain sciences: perils and prospects*, eds S. Rose and D. Rees, pp. 167 – 180. Cambridge University Press.
- Ferguson, T. S. (1973) A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1**(2), 209 – 230.
- Ferguson, T. S. (1974) Prior distributions on spaces of probability measures. *The Annals of Statistics* **2**(4), 615 – 629.
- Friedrich, J. and Paninski, L. (2016) Fast active set methods for online spike inference from calcium imaging. In *Advances In Neural Information Processing Systems*, eds D. Lee, M. Sugiyama, U. Luxburg, I. Guyon and R. Garnett, pp. 1984 – 1992.
- Friedrich, J., Zhou, P. and Paninski, L. (2017) Fast online deconvolution of calcium imaging data. *PLOS Computational Biology* **13**(3), 1 – 26.

- Frühwirth-Schnatter, S. and Malsiner-Walli, G. (2019) From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering. *Advances in Data Analysis and Classification* **13**, 33–64.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A. and Rubin, D. (2013) *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.
- Grienberger, C. and Konnerth, A. (2012) Imaging calcium in neurons. *Neuron* **73**(5), 862 – 885.
- Jewell, S. and Witten, D. (2018) Exact spike train inference via L0 optimization. *The Annals of Applied Statistics* **12**(4), 2457 – 2482.
- Jewell, S. W., Hocking, T. D., Fearnhead, P. and Witten, D. M. (2019) Fast nonconvex deconvolution of calcium imaging data. *Biostatistics* **21**(4), 709–726.
- Malsiner-Walli, G., Frühwirth-Schnatter, S. and Grün, B. (2016) Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and Computing* **26**, 303 – 324.
- McCullagh, P. and Yang, J. (2008) How many clusters? *Bayesian Analysis* **3**(1), 101 – 120.
- Miller, J. W. and Harrison, M. T. (2018) Mixture models with a prior on the number of components. *Journal of the American Statistical Association* **113**(521), 340–356.
- Mukamel, E. A., Nimmerjahn, A. and Schnitzer, M. J. (2009) Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* **63**(6), 747 – 760.
- Nobile, A. (2004) On the posterior distribution of the number of components in a finite mixture. *The Annals of Statistics* **32**(5), 2044 – 2073.
- Nobile, A. and Fearnside, A. (2007) Bayesian finite mixtures with an unknown number of components: the allocation sampler. *Statistics and Computing* **17**, 147 – 162.
- Parker, D. (2006) Complexities and uncertainties of neuronal network function. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **361**(1465), 81 – 99.
- Parker, D. (2010) Neuronal network analyses: premises, promises and uncertainties. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **365**(1551), 2315 – 2328.
- Pnevmatikakis, E., Merel, J., Pakman, A. and Paninski, L. (2013) Bayesian spike inference from calcium imaging data. In *In Signals, Systems and Computers*, pp. 349 – 353.
- Pnevmatikakis, E. A., Soudry, D., Gao, Y., Machado, T. A., Merel, J., Pfau, D., Reardon, T., Mu, Y., Lacefield, C., Yang, W., Ahrens, M., Bruno, R., Jessell, T. M., Peterka, D. S., Yuste, R. and Paninski, L. (2016) Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**(2), 285 – 299.

- Richardson, S. and 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.
- Sethuraman, J. (1994) A constructive definition of Dirichlet priors. *Statistica Sinica* **4**(2), 639–650.
- Vogelstein, J. T., Packer, A. M., Machado, T. A., Sippy, T., Babadi, B., Yuste, R. and Paninski, L. (2010) Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology* **104**(6), 3691–3704.