

# INFERRING STRUCTURED CONNECTIVITY FROM SPIKE TRAINS UNDER NEGATIVE-BINOMIAL GENERALIZED LINEAR MODELS

BY SCOTT W. LINDERMAN, RYAN P. ADAMS, AND JONATHAN W. PILLOW

*Harvard University, Harvard University, Princeton University*

**Abstract.** The steady expansion of neural recording capability provides exciting opportunities for discovering unexpected patterns and gaining new insights into neural computation. Realizing these gains requires flexible and accurate yet tractable statistical methods for extracting structure from large-scale neural recordings. Here we present a model for simultaneously recorded multi-neuron spike trains with negative binomial spiking and structured patterns of functional coupling between neurons. We use a generalized linear model (GLM) with negative-binomial observations to describe spike trains, which provides a flexible model for over-dispersed spike counts (i.e., responses with greater-than-Poisson variability), and introduce flexible priors over functional coupling kernels derived from sparse random network models. The coupling kernels capture dependencies between neurons by allowing spiking activity in each neuron to influence future spiking activity in its neighbors. However, these dependencies tend to be sparse, and to have additional structure that is not exploited by standard (e.g., group lasso) regularization methods.

For example, neurons may belong to different classes, as is often found in the retina, or they may be characterized by a small number of features, such as a preferred stimulus selectivity. These latent variables lend interpretability to otherwise incomprehensible data. To incorporate these concepts, we decompose the coupling kernels with a weighted network, and leverage latent variable models like the Erdős-Renyi model, stochastic block model, and the latent feature model as priors over the interactions.

To perform inference, we exploit recent innovations in negative binomial regression to perform efficient, fully-Bayesian sampling of the posterior distribution over parameters given the data. This provides access to the full posterior distribution over connectivity, and allows underlying network variables to be inferred alongside the low-dimensional latent variables of each neuron. We apply the model to neural data from primate retina and show that it recovers interpretable patterns of interaction between different cell types.

**1. Model.** We begin with vectors of simultaneously recorded spike counts,  $\{\mathbf{s}_n\}_{n=1}^N$ , from  $N$  neurons, where each vector  $\mathbf{s}_n \in \mathbb{N}_+^T$  represents  $T$  time bins. Typically, the bin size is taken to be small enough that only a few spikes may be observed, for example 10ms. We will model these spike counts as random draws from a negative binomial distribution,

$$(1) \quad s_{n,t} \sim \text{NB}(\xi, p_{n,t}) = \binom{s_{n,t} + \xi - 1}{s_{n,t}} (1 - p_{n,t})^\xi p_{n,t}^{s_{n,t}},$$

for  $\xi > 0$  and  $p_{n,t} \in (0, 1)$ . Letting  $\lambda_{n,t} = \frac{p_{n,t}}{1-p_{n,t}}$ , we have that  $\mathbb{E}[s_{n,t}] = \xi \lambda_{n,t}$  and  $\text{Var}[s_{n,t}] = \xi \lambda_{n,t} (1 + \lambda_{n,t})$ . We proceed by modeling  $\lambda_{n,t}$  with a log linear model. Let,

$$(2) \quad \log \boldsymbol{\lambda}_n \equiv \boldsymbol{\psi}_n \sim \mathcal{N}(\boldsymbol{\mu}_\psi, \sigma_\psi^2 \mathbf{I}), \quad \boldsymbol{\mu}_\psi = b_n \mathbf{1}_T + \sum_{n'=1}^n A_{n' \rightarrow n} \cdot \mathbf{X}_{n' \rightarrow n} \mathbf{w}_{n' \rightarrow n},$$

where  $b_n \sim \mathcal{N}(\mu_b, \sigma_b^2)$  is a constant bias for neuron  $n$ ;  $A_{n' \rightarrow n} \in \{0, 1\}$  is a binary variable indicating whether or not the spikes on neuron  $n'$  affect the firing of neuron  $n$ ;  $\mathbf{X}_{n' \rightarrow n} \in \mathbb{R}^{T \times B}$  is a matrix of given regressors, typically the filtered spike history; and  $\mathbf{w}_{n' \rightarrow n} \in \mathbb{R}^B$  is a vector of weights capturing the directed effect of neuron  $n'$  on neuron  $n$ . Together,  $A$  and  $\mathbf{w}$  represent a sparse, directed network of functional relationships.

This *spike-and-slab* model yields sparse parameter estimates and lends interpretability to an otherwise complex set of  $O(N^2)$  variables. Moreover, we often expect to find interpretable structure within the functional network. In this abstract we will focus on one particular type of structure where each neuron belongs to a latent class  $c_n \in [1, \dots, C]$ . These latent classes govern the sparsity of the functional network through the

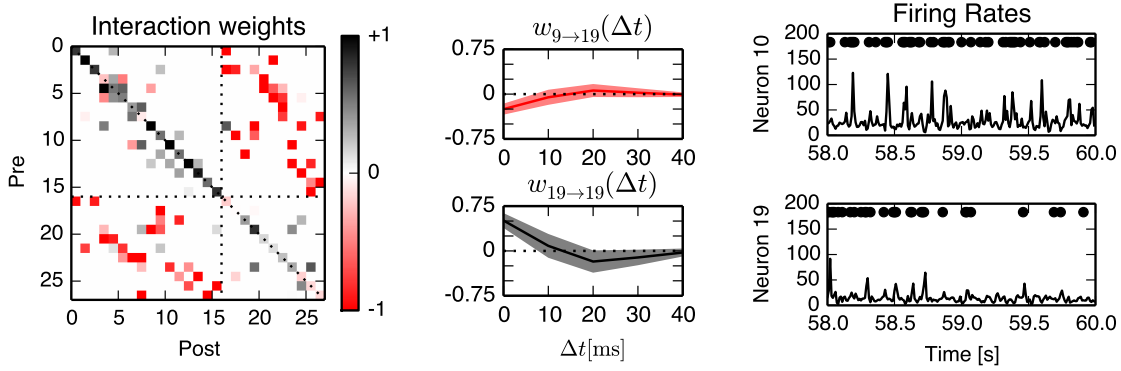


Fig 1: The negative binomial GLM applied to a population recording of retinal ganglion cells. Neurons 1–16 are “OFF” cells, and 17–27 are “ON” cells. The sparse network of inferred interactions is shown on the left, with canonical patterns of within-type excitation (black) and between-type inhibition (red). Each connection corresponds to a weighted impulse response, like those shown in the center. The sum of weighted impulse responses are passed through an exponential link function to determine the expected number of spikes per bin under the negative binomial model (right).

model  $A_{n' \rightarrow n} \sim \text{Bernoulli}(\rho_{c_{n'} \rightarrow c_n})$ , and the weights through  $\mathbf{w}_{n' \rightarrow n} \sim \mathcal{N}(\boldsymbol{\mu}_{c_{n'} \rightarrow c_n}, \boldsymbol{\Sigma}_{c_{n'} \rightarrow c_n})$ . For example, in the retina these classes could distinguish “ON” and “OFF” cells. In this “stochastic block model,” the classes are drawn from a categorical distribution with parameter  $\mathbf{p}$ . Conjugate Dirichlet and normal-inverse Wishart priors are placed over the parameters to complete the model.

**2. Bayesian Inference.** Pillow and Scott (2012) have derived a data augmentation scheme to efficiently perform Bayesian inference in negative binomial regression models. The fundamental property they exploit is that the likelihood as a function of  $\psi_{n,t} = \log \lambda_{n,t}$  can be written as,

$$(3) \quad p(s_{n,t} | \psi_{n,t}, \xi) \propto \frac{(e^{\psi_{n,t}})^{s_{n,t}}}{(1 + e^{\psi_{n,t}})^{s_{n,t} + \xi}} \propto \exp\left(\frac{s_{n,t} - \xi}{2} \psi_{n,t}\right) \mathbb{E}_{\omega_{n,t}} [\exp(-\omega_{n,t} \psi_{n,t}^2 / 2)],$$

where  $\omega_{n,t} \sim \text{PG}(s_{n,t} + \xi, 0)$  and PG denotes the Polya-Gamma distribution. Hence, if we augment our data with Polya-Gamma random variables  $\omega_{n,t}$ , then conditioned on these variables the log likelihood is a quadratic function of  $\psi_{n,t}$  and conjugate with a Gaussian prior. Similarly, the posterior distribution over  $\boldsymbol{\omega}_n$  factorizes into independent Polya-Gamma distributions that can be efficiently sampled. This data augmentation scheme renders our model conjugate and hence amenable to Gibbs sampling. The only exception is the spike-and-slab prior over  $A_{n' \rightarrow n}$  and  $\mathbf{w}_{n' \rightarrow n}$ . To improve the mixing of our Markov chain, we jointly update these parameters by marginalizing over the weights to sample  $A$  and then sampling  $\mathbf{w}$  given  $A$ . Since the weights and the likelihood are both Gaussian distributions, the marginal likelihood is also Gaussian and can be computed in closed form. Integrating out weights removes the conditional independence of the elements of  $\boldsymbol{\psi}_n$ , but the resulting marginal distribution is Gaussian with a low rank plus diagonal covariance matrix, and hence can be efficiently inverted with the matrix inversion lemma.

To demonstrate this approach, we have applied our negative binomial GLM to a population of primate retinal ganglion cells. With a simple Erdős-Renyi network prior, we recover a sparse pattern of connectivity that filters out the insignificant interactions and leaves an obvious pattern of mutual excitation among cells of the same type (diagonal blocks) and inhibition between cells of different types (off-diagonal blocks), as shown in Figure 1.

As we look toward increasingly large multiple spike train recordings, we plan to leverage many nice properties of this method. First, the data augmentation used to sample the negative binomial model is naively parallelizable. Moreover, given the conjugacy of this model, stochastic variational inference is an appealing alternative to Gibbs sampling that combines the benefits of Bayesian inference with the scalability required to discover structure in large scale recordings.

PILLOW, J. W. and SCOTT, J. (2012). Fully Bayesian inference for neural models with negative-binomial spiking. In *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, eds.) 1898–1906.