Discovering Latent Network Structure in Neural Spike Trains

Scott W. Linderman
Harvard University

February 12, 2015
UMass Amherst
Suppose you are studying a population of neurons...

How would you describe this activity?

$t = 0.0$
Suppose you are studying a population of neurons...
How would you describe this activity?

$t = 0.0$
Suppose you are studying a population of neurons...
How would you describe this activity?

$t = 0.0$
What if I showed you this?

Source: Jeremy Freeman
Latent Network Discovery

What if we cannot see the edges, only the activity of each node?
A Probabilistic Approach

\[ \Pr \left( \begin{array}{c}
\text{Random Network Model}
\end{array} \right) = \Pr \left( \begin{array}{c}
\text{Point Process Model}
\end{array} \right) \]
Agenda

1. Random network models

2. Mutually excitatory point processes
   a. Bayesian inference via auxiliary variable formulation

3. Mixed excitatory-inhibitory networks
   a. Polya-gamma augmentation

4. Extensions
A Probabilistic Approach

\[ \Pr(O, \mathcal{E}) = \Pr(\mathcal{E}) \times \Pr(O) \]

- **Random Network Model**
- **Point Process Model**
Erdös-Renyi Model

Most neurons do not interact with one another.

\[ A_{n,n'} \sim \text{Bern}(\rho) \]
\[ W_{n,n'} \sim p(W_{n,n'})^{A_{n,n'}} \delta_0^{1-A_{n,n'}} \]

Effectively \( L_0 \) regularization.
Latent Feature Models

Neurons have **latent features**; neurons that are nearby in feature space are more likely to interact.

\[
\ell_n \sim \mathcal{N}(\mu, \sigma^2)
\]

\[
A_{n,n'} \sim \text{Bern}(e^{-\|\ell_{n'} - \ell_n\|/\tau})
\]

\[
W_{n,n'} \sim p(W_{n,n'} | \ell_n, \ell_{n'}) A_{n,n'} \delta_{0}^{1-A_{n,n'}}
\]
Neurons have a **latent type** that governs how they interact.

\[
c_n \sim \text{Cat}(C) \\
A_{n,n'} \sim \text{Bern}(\rho_{c_n,c_{n'}}) \\
W_{n,n'} \sim p(W_{n,n'} \mid c_n, c_{n'}) A_{n,n'} \delta_0^{1-A_{n,n'}}
\]
Random Graph Models

- Huge assortment of extensions and variations:
  - Infinite Relational Model (Kemp et. al. 2006)
  - Mixed Membership Stochastic Block Model (Airoldi et. al., 2008)
  - Eigenmodel (Hoff 2007)
  - Random Function Model (Lloyd et. al. 2012)
1. Random network models
2. **Mutually excitatory point processes**
   a. Bayesian inference via auxiliary variable formulation
3. Mixed excitatory-inhibitory networks
   a. Polya-gamma augmentation
4. Extensions
Probabilistic Framework

Next, we need a model to relate the network to the observed spiking activity.
Preliminaries: Poisson processes

- Distribution over sets of spikes
- Rate function \( \lambda(t) \)
- Poisson-distributed number of events
- Non-overlapping time intervals are independent

**Poisson superposition principle**

The sum of independent Poisson processes with rates \( \lambda_1(t) \) and \( \lambda_2(t) \) is itself a Poisson process with rate:

\[
\lambda(t) = \lambda_1(t) + \lambda_2(t)
\]
Hawkes process

- Hawkes processes are multivariate point processes with conditionally Poisson dynamics.

- Hawkes processes introduce dependencies among spikes, according to an underlying network.
Hawkes process dynamics
What if some interactions are stronger than others?
Weighted Hawkes process dynamics

“weight” = area under curve
= expected # of events
Discrete Time Formulation

\[ s_{n,t} \sim \text{Poisson}(\lambda_{n,t}) \]
Formalizing this illustration

\[ \lambda_{t,n} \triangleq \lambda_n^{(0)} + \sum_{n' = 1}^{N} \sum_{t' = 1}^{t-1} s_{t',n'} \cdot h_{n' \rightarrow n}[t - t'], \]

Baseline activation

Directed, causal impulse response

\[ p(s \mid \lambda) = \prod_{t=1}^{T} \prod_{n=1}^{N} \text{Poisson}(s_{t,n} \mid \lambda_{t,n} \Delta t). \]
Incorporating the network

\[ h_{n' \rightarrow n}[t - t'] \triangleq W_{n' \rightarrow n} \cdot r_{n' \rightarrow n}[t - t'] \]
Impulse Response Model

Convex combination of normalized basis functions

\[ r_{n' \rightarrow n'}[t - t'] = \sum_b g_b^{(n',n)} \phi_b[t - t'] \]

\[ \phi_1[t - t'] \]
\[ \cdot \]
\[ \cdot \]
\[ \phi_B[t - t'] \]
Putting it all together

\[ \lambda_{t,n} = \lambda_{n}^{(0)} + \sum_{n'=1}^{N} \sum_{d=1}^{D} \sum_{b=1}^{B} s_{t-d,n'} W_{n'} \to n g_{b}^{(n',n)} \phi_{b}[d] \]

\[ = \lambda_{n}^{(0)} + \sum_{n'=1}^{N} \sum_{b=1}^{B} (s_{n'} \ast \phi_{b})[t] \cdot W_{n'} \to n g_{b}^{(n',n)} \]

\[ = \lambda_{n}^{(0)} + \sum_{n'=1}^{N} \sum_{b=1}^{B} \hat{s}_{t,n',b} W_{n'} \to n g_{b}^{(n',n)} \]
Agenda

1. Random network models
2. Mutually excitatory point processes
   a. Bayesian inference via auxiliary variable formulation
3. Mixed excitatory-inhibitory networks
   a. Polya-gamma augmentation
4. Extensions
Bayesian Inference

- Compute the posterior distribution over networks given the observed spike train.

- MAP estimation is a convex optimization problem in the “standard Hawkes” model, but not with network priors.

- MCMC and variational methods give us an approximation of the posterior, not just a point estimate.
Auxiliary “Parent” Variables

\[ \lambda_{t,n} = \lambda_{n}^{(0)} + \sum_{n'=1}^{N} \sum_{b=1}^{B} \hat{s}_{t,n',b} W_{n'\rightarrow n} g_{b}^{(n',n)} \]

\[ \text{or} \]

\[ \hat{s}_{t,n,b} W_{n'\rightarrow n} g_{b}^{(n',n)} \]

\[ \sim \frac{t}{n} \]

\[ \hat{s}_{t,n,b} W_{n'\rightarrow n} g_{b}^{(n',n)} \]
The parents are \textit{conditionally multinomial}:

\[ z_{t,n} \mid s_{t,n}, \lambda_n^{(0)}, W, g \sim \text{Mult}(s_{t,n}, p_{t,n}), \]

\[ p_{t,n} = \left[ \lambda_n^{(0)}, \ldots, \hat{s}_{t,n'}, b W_{n' \rightarrow n} g_{b}^{(n',n)}, \ldots \right] \cdot \frac{1}{\lambda_{t,n}}. \]

With the parents, the \textbf{likelihood factorizes}:

\[ p(z \mid \lambda) = \prod_{t=1}^{T} \prod_{n=1}^{N} \text{Pois}(z_{t,n}^{(0)} \mid \lambda_n^{(0)} \Delta t) \]

\[ \times \prod_{t=1}^{T} \prod_{n=1}^{N} \prod_{n'=1}^{N} \text{Pois}(z_{t,n'}^{(n',b)} \mid \hat{s}_{t,n'}, b W_{n' \rightarrow n} g_{b}^{(n',n)} \Delta t). \]
Due to **gamma-Poisson conjugacy**,

\[
\lambda_{n}^{(0)} \mid z \sim \text{Gamma}(\alpha \lambda + \sum_{t=1}^{T} z_{t,n}^{(0)}, \beta \lambda + T \Delta t)
\]

Similarly, the impulse responses have conjugate, **Dirichlet priors**:

\[
g^{(n',n)} \mid z, \gamma \sim \text{Dir} \left( \gamma^{(n',n)} \right),
\]

\[
\gamma_{b}^{(n',n)} = \gamma_{b} + \sum_{t=1}^{T} z_{t,n}^{(n',b)}.
\]
Gibbs Sampling the Weights

- The spike-and-slab model is not conjugate.

\[ p(W | A) = \text{Gam}(\kappa, \nu)^A \cdot \delta_0^{1-A} \]

- But smoothing the delta function to a mixture of gammas renders it conjugate.

- A and \( W \) are then updated jointly.
Gibbs Sampling the Network

- For example, in an SBM we can specify conjugate priors for the latent parameters:

\[ c_n \sim \mathcal{M}, \]
\[ A_{n' \rightarrow n} \sim \text{Bern}(p_{c_{n' \rightarrow c_n}}), \]
\[ W_{n' \rightarrow n} \sim \text{Gam}(\kappa, \nu_{c_{n' \rightarrow c_n}})^{A_{n' \rightarrow n}} \times \text{Gam}(\kappa_0, \nu_0)_{1-A_{n' \rightarrow n}} \]
Stochastic Variational Inference

- The parent variables are **conditionally independent** given the rate, and the model is **fully conjugate**.

- SVI on mini-batches of data follows directly.

\[
\begin{array}{cccccccccc}
1 & 1 & & & & & & & & \\
& 1 & & & & & & & & \\
1 & & & & & & & & & \\
& 1 & 1 & & & & & & & \\
1 & 1 & 1 & & & & & & & \\
1 & 1 & 1 & 1 & & & & & & \\
1 & 1 & 1 & 1 & 1 & & & & & \\
1 & 1 & 1 & 1 & 1 & 1 & & & & \\
& 1 & 1 & 1 & 1 & 1 & 1 & & & \\
\end{array}
\]
First, we test with ten **sparse** networks of 30 nodes each, **simulated from our model.**
Inference Algorithm Comparison

Next, we compared predictive likelihood as a function of wall-clock time.

\[ N = 50, \; T = 10^4 \]

\[ N = 50, \; T = 10^5 \]
Agenda

1. Random network models
2. Mutually excitatory point processes
   a. Bayesian inference via auxiliary variable formulation
3. Mixed excitatory-inhibitory networks
   a. Polya-gamma augmentation
4. Extensions
Mixed Networks

What if weights can be negative (inhibitory) as well?

$W > 0$

$W < 0$
Probabilistic Framework

To handle inhibition, we swap in a new point process model.
Mixed Network Dynamics

"Activation"
Mixed Network Dynamics
Mixed Network Dynamics
Mixed Network Dynamics
Mixed Network Dynamics
Mixed Network Dynamics
Mixed Network Dynamics
Mixed Network Dynamics
Formalizing this illustration

Sum over preceding spikes

\[ \psi_{t,n} \triangleq b_n + \sum_{n'=1}^{N} \sum_{t'=1}^{t-1} s_{t',n'} \cdot h_{n' \rightarrow n}[t - t'] \]

Baseline activation

Directed, causal impulse response

\[ s_{t,n} \sim p(s_{t,n} \mid \psi_{t,n}) \]

Spike count model

“Point process generalized linear model”

- Truccolo et al, 2003
- Paninski, 2004
- Pillow et al, 2008
## Comparison of Spike Count Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s_{n,t} \sim \text{Bernoulli}(\sigma(\psi_{n,t})))</td>
<td>- Appropriate for small time bins.</td>
<td>- Small time bins implies many terms in likelihood.</td>
</tr>
<tr>
<td>(\lambda_{n,t} = g(\psi_{n,t}))</td>
<td>- Essentially logistic regression.</td>
<td></td>
</tr>
<tr>
<td>(s_{n,t} \sim \text{Poisson}(\lambda_{n,t}))</td>
<td>- For simple models, MAP estimation is easy.</td>
<td>- Bayesian inference is challenging.</td>
</tr>
<tr>
<td>(s_{n,t} \sim \text{Neg.Bin.}(\xi, \sigma(\psi_{n,t})))</td>
<td>- Over-dispersed.</td>
<td>- Fixed mean/variance relationship.</td>
</tr>
<tr>
<td></td>
<td>- Polya-Gamma trick renders the model conjugate.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Could be too flexible for small datasets.</td>
<td>- Less biophysically plausible.</td>
</tr>
</tbody>
</table>
1. Random network models
2. Mutually excitatory point processes
   a. Bayesian inference via auxiliary variable formulation
3. Mixed excitatory-inhibitory networks
   a. Polya-gamma augmentation
4. Extensions
This form appears in both the Bernoulli and negative binomial likelihoods.

Polson et al. (2011) introduced the Polya-gamma distribution with the property that,

\[
\frac{(e^{\psi t,n})^{s t,n}}{(1 + e^{\psi t,n})^{s t,n + \xi}}
\]

Conditioned on \( \omega_{n,t} \), the log likelihood is quadratic in \( \psi_{n,t} \)!

\[
\left(\frac{e^{\psi t,n}}{1 + e^{\psi t,n}}\right)^a \left(1 + e^{\psi t,n}\right)^b \equiv 2^{-b} e^{\kappa \psi t,n} \int_0^\infty e^{-\omega t,n \psi^2 t,n/2} \rho(\omega t,n) d\omega t,n
\]

\( \omega_{t,n} \sim \text{PG}(b, 0), \kappa = a - b/2 \)
Linear Gaussian Activation

We use a linear Gaussian model to capitalize on the conjugacy afforded by the Polya-gamma augmentation.

\[
\psi_{t,n} \overset{\Delta}{=} b_n + \sum_{n'=1}^{N} \sum_{t'=1}^{t-1} s_{t',n'} W_{n' \rightarrow n}^{(j)} \phi_j [t - t']
\]

\[
= b_n + \sum_{n'=1}^{N} \sum_{j=1}^{J} \hat{s}_{t,n',j} W_{n' \rightarrow n}^{(j)}
\]

\[
= b_n + \sum_{n'=1}^{N} \hat{s}_{t,n'}^T W_{n' \rightarrow n}
\]

\[
b_n \sim \mathcal{N}(\mu_b, \sigma_b^2)
\]

\[
W \sim \mathcal{N}(\mu_W, \Sigma_W)^A \delta_0^{1-A}
\]
Polya-gamma augmentation

- Conditioned on the activation, the auxiliary variables are still **Polya-gamma distributed**.

\[ \omega_{n,t} \mid \psi_{n,t} \sim \text{PG}(s_{n,t} + \xi, \psi_{n,t}) \]

- Efficient samplers exist for the PG distribution.
- The Polya-gamma was designed in order to perform efficient Bayesian inference in logistic and binomial models.
  - Polson, Scott, and Windle (2013)
  - Zhou, Li, Dunson, and Carin (2013)
  - Pillow and Scott (2013)
Augmented Network Model

We again introduce Polya-gamma auxiliary variables to render the model conjugate.
Retinal Ganglion Cell Network Model

Each neuron has a **latent type** and **location**.

\[ \ell_n \in [0, 10]^2 \quad c_n \in \{1, \ldots, C\} \]

**Type** governs **weight distribution**.

\[ W_{n' \rightarrow n} \sim \mathcal{N}(\mu_{c_{n' \rightarrow n}}, \sigma_{c_{n' \rightarrow n}}^2) \]

**Location** governs **probability of interaction**.

\[ A_{n' \rightarrow n} \sim \text{Bern.}(e^{-\|\ell_{n'} - \ell_n\|^2/\tau}) \]
Retinal Ganglion Cells

Like cells excite and opposite inhibit in a spatially localized manner.
Agenda

1. Random network models
2. Mutually excitatory point processes
   a. Bayesian inference via auxiliary variable formulation
3. Mixed excitatory-inhibitory networks
   a. Polya-gamma augmentation
4. Extensions
Example 2: Hidden States of the Hippocampus

Joint work with Matt Johnson, Matthew Wilson, and Zhe Chen in submission
Switching Network Models

Spike train

Latent state

Transition Matrix

Firing rates

$S_t$

$S_{t+1}$
Switching Network Models

$z_t \in \{1, \ldots, K\}$

$z_1 \rightarrow \cdots \rightarrow z_{t-1} \rightarrow z_t \rightarrow \cdots \rightarrow z_T$

$\omega_1 \rightarrow \cdots \rightarrow \omega_{t-1} \rightarrow \omega_t \rightarrow \cdots \rightarrow \omega_T$

$s_1 \rightarrow \cdots \rightarrow s_{t-1} \rightarrow s_t \rightarrow \cdots \rightarrow s_T$
Application to hippocampal spike trains

Empirical Location Distribution

Location Trajectory

All states

State 1 (3.9%)

State 2 (3.6%)

State 3 (3.1%)

\[
p(\ell) \text{ [m}^2\text{]}\]
Dynamic networks and synaptic plasticity

Joint work with Chris Stock and Ryan Adams, NIPS 2014
Sparse time-varying networks

Similar to the switching GLM, but now the network is a **dynamical system**.
Time-varying Weights

\[ h_{n'\rightarrow n}(\Delta t, t) \equiv A_{n'\rightarrow n} W_{n'\rightarrow n}(t) r_{n'\rightarrow n}(\Delta t). \]
Modeling learning rules

Consider the standard additive STDP rule

\[
\frac{dW_{n' \to n}(t)}{dt} = s_n(t) \sum_{s_{n',m} < t} A_+ e^{(t-s_{n',m})/\tau_+} - s_{n'}(t) \sum_{s_{n,m} < t} A_- e^{(t-s_{n,m})/\tau_-}
\]

or the multiplicative version

\[
\frac{dW_{n' \to n}(t)}{dt} = s_n(t) \sum_{s_{n',m} < t} l_+(t-s_{n',m})(W_{\text{max}} - W(t)) \ldots
\]
Inference with Particle MCMC

- **Idea:** represent $p(W_t)$ as a set of weighted particles

- Propagate particles forward according to learning rule

- Reweight according to how well they match the spike train
Comparing learning rules

First, we tested on synthetic data from a GLM
Themes

• **Bayesian modeling:** combine top down intuition with bottom up evidence from data.

• **Modularity:** Separating network and observation models allows for flexibility and reuse.

• **Interpretability:** latent variables of the network should guide hypothesis generation.

• **Efficiency:** Auxiliary variables can make inference simple and scalable.
Acknowledgements

Chris Stock
Matt Johnson
Ryan Adams
Jonathan Pillow
Sage Chen
Matthew Wilson
Leslie Valiant

Harvard Intelligent Probabilistic Systems (HIPS) Group

NDSEG
Center for Brains Mind and Machines
Questions?

Thank you!

slinderman@seas.harvard.edu

http://people.seas.harvard.edu/~slinderman

https://github.com/slinderman